НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова
праця на правах рукопису

СОКОЛЬСЬКИЙ СЕРГІЙ ОЛЕГОВИЧ

УДК 681.884:004.93

ДИСЕРТАЦІЯ
АКУСТИЧНЕ ВИЯВЛЕННЯ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ

172 Телекомунікації та радіотехніка
17 Електроніка та телекомунікації

Подається на здобуття наукового ступеня доктор філософії
Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело
_____ С.О. Сокольський

Науковий керівник: Мовчанюк Андрій Валерійович, кандидат технічних наук, доцент

Київ – 2023

АНОТАЦІЯ

Сокольський С.О. Акустичне виявлення безпілотних літальних апаратів. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 172 «Телекомунікації та радіотехніка». – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» МОН України, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського» МОН України, Київ, 2023.

У дисертаційній роботі вирішено актуальне ***науково-технічне завдання*** — розроблення пристрою з одним вхідним сенсором-мікрофоном для виявлення малих безпілотних літальних апаратів на основі аналізу акустичних сигнатур. Розроблена модель системи дає змогу підвищити ефективність виявлення дронів за рахунок використання нейронної мережі для обробки акустичних сигналів.

Основний зміст дисертаційної дослідницької роботи наведено в чотирьох розділах, де представлено та обґрунтовано основний результат.

У вступній частині підтверджена актуальність роботи, сформульована мета, завдання та методи дослідження, надана інформація про наукову новизну, а також практичне значення отриманих результатів.

У першому розділі був проведений порівняльний та критичний аналіз потенційних можливостей основних методів виявлення малих безпілотних летальних апаратів (МЛА): оптичного, радіолокаційного та акустичного. Кожним із методів розраховано теоретичну максимальну дальність виявлення безпілотника моделі *«DJI Mavic 3»* який зараз активно використовується військовими для розвідки. Встановлено, що оптичний метод передбачає використання камер високої розділової здатності у світлу пору дня, але він занадто залежний від природних чинників навколишнього середовища, наприклад, дощу або туману. У темну пору доби зазвичай використовують інфрачервоні оптичні пристрої. Основними недоліками такого підходу є невисока максимальна дальність виявлення цілі та поглинання атмосферою ІЧ

випромінювання, крім «вікон» із межами довжин хвиль 3 – 4 та 8 – 12 мм. У результаті наведених розрахунків встановлено, що дальність визначення цілі оптичним методом складає біля 250 м, а з використанням ІЧ приймача — 79,5 м. Зазначено, що радіолокаційні методи поділяються на активні та пасивні. Радіолокаційні системи можуть працювати в будь-яку частину доби та дають змогу виявляти МЛА на відстанях до кількох кілометрів (9,8 км). У результаті аналізу спектру радіочастотного сигналу можна отримати найбільш детальну інформацію про ціль. Основним недоліком радіолокаційних методів є те, що всі РЛС не працюють у ближній зоні. Встановлено, що акустичні сенсори дають змогу ефективно виявляти безпілотні летальні апарати, незважаючи на рельєф оточуючого середовища, але результати залежать від наявних акустичних фонових шумів та завад. Ефективна дальність виявлення дрону склала біля 78 м. У висновках зазначено, що оскільки МЛА мають невисокий рівень ІЧ випромінювання та радіолокаційних сигнатур, то актуальним стає питання розроблення детектора із використанням акустичного методу виявлення дронів та великою максимальною відстанню до цілі.

У другому розділі розглянуто реалізацію простого та дешевого для конструювання акустичного детектора з одним мікрофоном. Він дає змогу виявляти акустичні коливання, що видають двигуни та лопоті дрону, і який потім можна проаналізувати для визначення місцезнаходження та траєкторію руху МЛА. Для реєстрації детектором звукових хвиль було запропоновано використовувати електретний мікрофон CMA-4544PF-W із поролоновою вітрозахисною насадкою, а для забезпечення широкого динамічного діапазону та захисту детектора від перевантаження — підсилювач із системою автоматичного регулювання підсилення зі зворотним регулюванням на базі інтегральної мікросхеми MAX9814. Для подальшої цифрової обробки сигналів на ПК, буде використовуватись аналого-цифровий перетворювач із частотою дискретизації 48 кГц і розрядністю 16 біт, та зовнішня звукова карта CM6206 із лінійним входом. За результатом тестових випробувань була створена база аудіофайлів шуму квадрокоптера «*DJI Mavic 3*». Спектри отриманих аудіозаписів

випромінювання квадрокоптера надають змогу визначити основні частоти тонів, кількість яких збігається з кількістю електромоторів та є важливими ознаками під час розроблення алгоритму класифікації безпілотників.

Третій розділ описує процес розроблення алгоритму для ефективного виявлення та класифікації аудіосигналів МЛА із використанням згорткової нейронної мережі глибокого навчання, побудови архітектури та теоретичного оцінювання ефективності її роботи. Перед подачею набору аудіозаписів дронів на вхід нейронної мережі, підвищили їхню якість, застосувавши нормалізацію, Вінеровську фільтрацію, сегментацію — поділили аудіо на кадри тривалістю 25 мс із перекриттям 50% та віконне керування за допомогою вікна Хеммінга, оскільки в завданні оброблення аудіосигналів важливіша точність у часовій області. Отримані дані розділили в співвідношенні 60/20/20 на три набори: для навчання, валідації та тестування. Далі представили дані спрощеним набором ознак, визначивши з кожного кадру оброблених аудіосигналів мел-спектрограми, для фіксації часових та спектральних характеристик. Діапазон частот аналізу становить межі робочих частот моделі мікрофону (20 Гц — 20 кГц), частотна розділова здатність 50 Гц, а кількість робочих мел-смуг дорівнює 30. Використовуючи навчальні дані та витягнуті ознаки аудіосигналів, розробили архітектуру нейронної мережі для досліджень роботи алгоритму виявлення та класифікації дронів. Вона складається із 10 пар шарів згортки, ReLU, пакетної нормалізації та шарів максимального пулингу. Їхня кількість визначається розміром вікна об'єднання вздовж часового виміру. Далі ідуть шари згладжування, відсікання, повнозв'язний та Softmax. Для нормалізації вихідних даних і отримання фінальних ймовірностей застосовується шар класифікації. У якості оптимізатора для навчання моделі обрано Adam. Відповідно до характеристик набору даних, початкова швидкість навчання дорівнює 0.001, а після проходження 75% епох поступово зменшується в 10 разів, для покращення збіжності. Точність розпізнавання вхідних даних складає 99 %, оцінка F1 — 0.93, що вказує на високий рівень загальної продуктивності архітектури.

У четвертому розділі проведено практичне тестування роботи готової моделі згорткової нейронної мережі. Ефективна дистанція виявлення МЛА алгоритмом дорівнює 200 м, що за методом експертних оцінок, значно перевершує можливості людського слуху з максимальною відстанню до 120 м. Підтверджено, що розроблений алгоритм може замінити людський персонал в операціях спостереження за дронами, пропонуючи підвищену ефективність і надійність під час захисту громадян, об’єктів критичної інфраструктури та національної оборони. Також надані рекомендації для подальшого покращення продуктивності та адаптивності алгоритму: використання різноманітних наборів даних, мультимодальне зондування, постійне вдосконалення, співпраця та використання передових методів машинного навчання.

Наукова новизна отриманих результатів полягає в тому, що:

1) Вперше запропоновано математичну модель ідентифікатора МЛА, наукова новизна якої полягає в тому, що вона ґрунтується на використанні бази аудіозаписів акустичних сигнатур дронів та надає можливість зменшити похибку під час класифікації їхньої моделі.

2) Вперше розроблено метод цифрової обробки акустичних сигналів дронів, наукова новизна якого полягає в тому, що він ґрунтується на теорії штучного інтелекту та використовує згорткову нейронну мережу глибокого навчання з оптимізатором Adam для підвищення швидкості та ефективності ідентифікації типу малих безпілотних літальних апаратів.

3) Удосконалено методику розрахунку кількості мел-фільтрів та їхнього частотного діапазону, що, на відміну від існуючих, враховує бажану частотну розділову здатність та дає змогу підвищити швидкість та точність отримання основних представлень аудіосигналу МЛА під час використання техніки зважування коефіцієнтів ШПФ за допомогою банків мел-фільтрів.

Практичне значення отриманих результатів полягає в тому, що:

1) На основі розробленої математичної моделі, створено базу аудіосигналів малих безпілотних літальних апаратів моделей «*Mavic 2 Pro*»,

«*Mavic 3*» та «*FPV*», що використовуються військовими в розвідувальних операціях, а також цивільного квадрокоптера «*Feilun Fx137*».

2) На основі розробленого методу та методики, написано програмне забезпечення для швидкого та ефективного оброблення аудіосигналів малих безпілотних літальних апаратів із використанням згорткової нейронної мережі глибокого навчання з оптимізатором Adam.

3) Створено акустичний детектор малих безпілотних летальних апаратів з ефективною відстанню виявлення об’єктів до 200 метрів.

4) Надано відповідні універсальні рекомендації для подальшого вдосконалення системи детекції малих безпілотних літальних апаратів.

Практичний результат дисертаційного дослідження може бути використаний під час побудови ефективних антидронних систем для оборони громадян та об’єктів критичної інфраструктури.

Частину результатів дисертаційної роботи отримано та впроваджено на підприємстві ТОВ «ТІНСТРУМ» за договором № Дндч/0201.01/2100.02/47/2023.

Ключові слова: дрон, малий безпілотний літальний апарат, БПЛА, виявлення цілей, акустична локація, контрольована зона, реєстратор сигналів, виявлення сигналів, спектр, оброблення сигналу, машинне навчання, нейронна мережа, глибоке навчання.

ABSTRACT

Serhii Sokolskyi. Acoustic detection of unmanned aerial vehicles. – Qualifying scientific work on the rights of the manuscript.

Thesis for the degree of Philosophy Doctor, in specialty 172 “Telecommunications and Radio Engineering”. – National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” of the Ministry of Education and Science of Ukraine, National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute” Ministry of Education and Science of Ukraine, Kyiv, 2023.

The thesis solves an urgent ***scientific*** and ***technical issue*** such as the development of a device with a single input sensor-microphone for detecting small unmanned aerial vehicles based on the analysis of acoustic signatures. The developed model of the system makes it possible to increase the efficiency of drone detection by using a neural network for processing acoustic signals.

The research aims to develop a prototype of the acoustic detector of small unmanned aerial vehicles, as well as to create a software algorithm for the effective processing of acoustic signals registered by the sensor for the presence of drone signatures.

The main content of the thesis is presented in four sections, where the main result is given and substantiated.

The introductory part confirms the relevance of the work, formulates the aim, tasks, and methods of research, also provides information about scientific novelty, as well as the practical significance of the obtained results.

In the first section, a comparative and critical analysis of the potential possibilities of the main methods of detecting small unmanned aerial vehicles, such as optical, radar and acoustic methods, is carried out. Each of the methods calculates the theoretical maximum detection range of the DJI Mavic 3 drone, which is currently actively used by the military for reconnaissance. It is found that the optical method involves the use of high-resolution cameras in daylight, but is too dependent on natural environmental factors such as rain or fog. In the dark, infrared optical devices are

usually used to detect drones. The main disadvantages of this approach are the low maximum target detection range and atmospheric absorption of IR radiation, except for “windows” with wavelength limits of 3 - 4 and 8 - 12 mm. As a result of the above calculations, it is determined that the target detection range by the optical method is about 250 m, and with the use of an IR receiver, it is 79.5 m. It is noted that radar methods are divided into active and passive. Radar systems can work at any time of the day and make it possible to detect UAVs at distances of up to several kilometers (9.8 km). By analyzing the spectrum of the RF signal, we can get the most detailed information about the target. The main disadvantage of radar methods is that all radars do not work in the near-field zone. It determines that acoustic sensors allow efficient detection of small unmanned aerial vehicles, despite the topography of the environment, but the results depend on the available acoustic background noise and interference. The effective target detection range is about 78 meters. The conclusions state that since UAVs have a low level of IR radiation and radar signatures, the issue of developing a detector using the acoustic method of drone detection and a large maximum distance to the target becomes relevant.

In the second section, we consider the implementation of a simple and cheap-to-construct acoustic detector with a single microphone. It allows detection of the sound emitted by the drone's engines and blades, which can then be analyzed to determine the location and trajectory of the UAV. It is proposed to use a CMA-4544PF-W electret microphone with a foam wind protection nozzle for the registration of sound waves by the detector, and to ensure a wide dynamic range and protect the detector from overload, an amplifier with an automatic gain adjustment system with reverse adjustment based on the MAX9814 integrated circuit. For further digital signal processing on a PC, an analog-to-digital converter with a sampling rate of 48 kHz and 16-bit resolution will be used, as well as an external CM6206 sound card with a line input. Based on the test results, a database of audio files of the noise of the DJI Mavic 3 quadcopter, which is now actively used by the military for reconnaissance, is created. The spectra of the received audio recordings of the radiation of the quadcopter make it possible to determine the main tone frequencies, the number of which coincides with

the number of electric motors and are important features in the development of the algorithm for the classification of drones.

The third section describes the process of developing an algorithm for efficient detection and classification of UAV audio signals using a convolutional deep learning neural network, building the architecture, and theoretically evaluating its performance. Before submitting the drone audio dataset to the neural network, the quality of the audio recordings is improved by applying normalization, Wiener filtering, and segmentation. The audio is segmented into frames with a duration of 25 ms and a 50% overlap, applying Hamming windowing for better accuracy in the time domain, as temporal precision is crucial in audio signal processing. The obtained data are divided in the ratio 60/20/20 into three sets: for training, validation, and testing. Next, the data is represented by a simplified set of features, extracting mel-spectrograms from each frame of the processed audio signals to capture their temporal and spectral characteristics. The frequency range of the analysis corresponds to the working frequency limits of the microphone model (20 Hz – 20 kHz), with a frequency resolution of 50 Hz and 30 working mel frequency bands. Using the training data and the extracted audio signals, a neural network architecture is developed for researching the operation of the drone detection and classification algorithm. It consists of 10 pairs of convolution layers, ReLU, batch normalization, and maximum pooling layers. Their number is determined by the size of the merging window along the time dimension. This is followed by flattening, dropout, fully connected, and Softmax layers. A classification layer is used to normalize the initial data and obtain the final probabilities. The Adam optimizer is chosen for model training. According to the characteristics of the data set, the initial learning rate is 0.001, and after passing 75% of the epochs, it is gradually reduced by a factor of 10 to improve the convergence. The accuracy of the input data recognition reaches 99%, and the F1 score of the trained model is 0.93, indicating a high level of overall architecture performance.

In the fourth section, practical testing of the finished convolutional neural network model is carried out. The effective distance of detection of UAV by the algorithm is 200 m, which, according to the method of expert evaluation, significantly

exceeds the capabilities of human hearing with a maximum distance of 120 m. It confirmes that the developed algorithm can replace human personnel in drone surveillance operations, offering increased efficiency and reliability when protecting citizens, critical infrastructure and national defense. Recommendations for further improving the performance and adaptability of the algorithm are also provided: the use of diverse datasets, multimodal sensing, continuous improvement, collaboration, and the use of advanced machine learning techniques.

The scientific novelty of the obtained results is that:

1) For the first time, a mathematical model of the UAV identifier is proposed. The scientific novelty of this model is that it is based on the use of a database of audio recordings of acoustic signatures of drones and makes it possible to reduce the error in classifying their model.

2) For the first time, a method of digital processing of acoustic signals of drones is developed, the scientific novelty of which is that it is based on the theory of artificial intelligence and uses a convolutional neural network of deep learning with the Adam optimizer to increase the speed and efficiency of the identification of the type of small unmanned aerial vehicles.

3) The method of calculating the number of mel-filters and their frequency range was improved, which, unlike the existing ones, takes into account the desired frequency resolution and makes it possible to increase the speed and accuracy of obtaining the main representations of the UAV audio signal when using the technique of weighting the FFT coefficients using banks of mel- filters.

The practical significance of the obtained results is that:

1) On the basis of the developed mathematical model, a database of audio signals of small unmanned aerial vehicles of the models "Mavic 2 Pro", "Mavic 3" and "FPV", used by the military in reconnaissance operations, as well as the civilian quadcopter "Feilun Fx137", was created.

2) On the basis of the developed method and methodology, software for fast and efficient processing of audio signals of small unmanned aerial vehicles using a convolutional neural network of deep learning with the Adam optimizer is written.

3) An acoustic detector of small unmanned aerial vehicles with an effective object detection distance of up to 200 meters is created.

4) Relevant universal recommendations for further improvement of the small unmanned aerial vehicle detection system were provided.

The practical result of the thesis research can be used during the construction of effective anti-drone systems for the defense of citizens and critical infrastructure.

Part of the results of the dissertation were obtained and implemented at TINSTRUM LLC under contract № Дндч /0201.01/2100.02/47/2023.

Keywords: drone, small unmanned aerial vehicle, UAV, target detection, acoustic location, controlled area, signal recorder, signal detection, spectrum, signal processing, machine learning, neural network, deep learning.

Список публікацій здобувача, в яких опубліковані основні наукові результати дисертації

1. С.О. Сокольський та А.В. Мовчанюк, «Огляд методів виявлення та локалізації малих безпілотних літальних апаратів», *Вісник НТУУ "КПІ". Серія Радіотехніка. Радіоапаратобудування,* №87, с. 46-55, Грудень. 2021. doi:10.20535/RADAP.2021.87.46-55 (фахове видання категорії А).

2. С.О. Сокольський та А.В. Мовчанюк, «Електроакустичний тракт детектора для виявлення малих безпілотних літальних апаратів», *Вісник Вінницького політехнічного інституту*, № 2, с. 135-144, Травень. 2023. doi: 10.31649/1997-9266-2023-167-2-135-144 (фахове видання категорії Б).

3. С.О. Сокольський та А.В. Мовчанюк, «Алгоритм оброблення аудіосигналів із використанням методу машинного навчання», *Вісник НТУУ "КПІ". Серія Радіотехніка. Радіоапаратобудування,* №93, с. 39-51, Вересень. 2023. doi:10.20535/RADAP.2023.93.39-51 (фахове видання категорії А).

4. С.О. Григораш (Сокольський) та А.В. Мовчанюк, «Нові виміри застосування дронів», на III *Міжнародній науково-практичній конференції "Філософія і наукова-технічна творчість у хронотопі технічного університету"*, Київ, Україна, 2020, с. 92-97 (матеріали конференції)

5. С.О. Сокольський та А.В. Мовчанюк, «Огляд методів виявлення і локалізації МЛА», на *X Міжнародній науково-технічній конференції «РТПСАС-2021*», Київ, Україна, 2021, с. 48-50 (матеріали конференції)

6. С.О. Сокольський та А.В. Мовчанюк, «Електроакустичний тракт детектора МЛА», на *XI Міжнародній науково-технічній конференції «РТПСАС-2022*», Київ, Україна, 2022, с. 88-90 (матеріали конференції)

Зміст

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ 15

ВСТУП 16

РОЗДІЛ 1. АНАЛІЗ ОСНОВНИХ МЕТОДІВ ВИЯВЛЕННЯ ТА ЛОКАЛІЗАЦІЇ МАЛИХ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ 21

1.1. Огляд завдання виявлення та локалізації малих безпілотних літальних апаратів 21

1.2. Методи виявлення і локалізації МЛА 23

1.2.1. Оптичний метод 23

1.2.2. Радіолокаційний метод 33

1.2.3. Акустичний метод 36

1.3. Постановка завдання дослідження 40

Висновки до розділу 1 42

РОЗДІЛ 2. РОЗРОБКА АКУСТИЧНОГО ТРАКТУ ДЕТЕКТОРА МАЛИХ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ 43

2.1. Вибір акустичного сенсора-мікрофона 44

2.2. Проектування підсилювача до мікрофона. Розподіл фільтрації між акустичним трактом та підсилювачем. 61

2.3. АЦП. Дискретизація сигналу 71

2.4. Формування бази сигналів МЛА 80

2.4.1. Опис вимірювальної установки 80

2.4.2. Умови проведення експерименту 81

2.4.3. Результати експерименту 81

Висновки до розділу 2 84

РОЗДІЛ 3. РОЗРОБКА АЛГОРИТМУ ДЛЯ ОБРОБЛЕННЯ АУДІОСИГНАЛІВ ІЗ ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ 85

3.1. Створення набору аудіоданих моделей МЛА 87

3.2. Попередня обробка вхідних аудіосигналів МЛА 88

3.3. Виділення ознак з вхідних аудіосигналів МЛА 93

3.4. Підготовка наборів навчальних даних 98

3.5. Створення та навчання моделі нейронної мережі 99

3.5.1. Архітектура моделі згорткової нейронної мережі 109

3.5.2. Навчання згорткової нейронної мережі 120

3.6. Оцінка та оптимізація розробленої моделі нейронної мережі 128

Висновки до розділу 3 132

РОЗДІЛ 4. РЕЗУЛЬТАТИ РОБОТИ СИСТЕМИ ВИЯВЛЕННЯ МАЛИХ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ 133

4.1. Тестування ефективності роботи системи виявлення МЛА 133

4.2. Рекомендації для підвищення ефективності роботи системи виявлення МЛА 148

Висновки до розділу 4 150

ЗАГАЛЬНІ ВИСНОВКИ 151

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 153

ДОДАТОК А 159

ДОДАТОК Б 161

ДОДАТОК В 162

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

АЧХ — амплітудно-частотна характеристика;

АЦП — аналого-цифровий перетворювач;

ІЧ — інфрачервоний;

МЛА — малі безпілотні літальні апарати;

НМ — нейронна мережа;

ПЗ — програмне забезпечення;

ПК — персональний комп’ютер;

ПФ — передавальна функція;

С/Ш — співвідношення сигнал / шум;

ФВЧ — фільтр високих частот;

ФНЧ — фільтр низьких частот;

ШПФ — швидке перетворення Фур’є;

ВСТУП

Актуальність теми. XXI століття — епоха активного розвитку робототехніки та штучного інтелекту. Вінцем її витвору можна за правом вважати малі безпілотні літальні апарати або дрони. Через свої габаритні розміри, ці компактні, маневрені та універсальні літальні апарати в останні роки почали використовуватися майже у всіх сферах діяльності для полегшення життя людей: від військової розвідки, патрулювання кордонів до аерофотозйомки, сільськогосподарських робіт та доставки вантажів у важкодоступні місця.

Нажаль зараз усе частіше застосування дронів сприяє незаконній діяльності: несанкціонованому спостереженню за критично важливими об’єктами інфраструктури, доставці контрабанди або бойових снарядів, що створюють величезну загрозу громадській та національній безпеці. У світлі цього постає важливе завдання ефективного та швидкого виявлення малих безпілотних літальних апаратів, що дасть можливість своєчасного розгортання контрзаходів для усунення потенційних негативних наслідків.

Детектування МЛА ускладнюють особливості траєкторії їхнього руху, а також малі габаритні розміри, що роблять такі дрони малопомітними в електромагнітному діапазоні довжин хвиль. Найбільш перспективним для цього завдання виглядає акустичний діапазон, оскільки шум від лопатей двигунів безпілотних літальних апаратів є серйозною демаскуючою ознакою, а акустичні хвилі прекрасно розповсюджуються в повітрі незважаючи на складність рельєфу та наявність зелених насаджень.

Аналізуючи ці сигнатури акустичними сенсорами, можна відрізнити звуки дронів від фонового шуму. Такий підхід передбачає використання мікрофонів для виявлення звуку та включає методи обробки аудіосигналів: спектральний аналіз, вилучення ознак і розпізнавання образів. Акустичні детектори такого типу складаються з багатоелементного масиву високоякісних мікрофонів, а також потребують недешеву апаратну базу та спеціальне програмне забезпечення.

Однак багато громадських місць та об'єктів не мають достатнього бюджету на придбання складних систем виявлення безпілотних літальних апаратів, які є високовартісними [1]. Часто на таких об'єктах просто хочуть знати, чи присутній дрон, і тому не потребують складних систем. Тому актуальним ***науково-технічним завданням*** цієї роботи є розроблення пристрою з одним вхідним сенсором-мікрофоном для ефективного виявлення малих безпілотних літальних апаратів на основі аналізу акустичних сигнатур.

Практичний результат дисертаційного дослідження може бути використаний під час побудови ефективних антидронних систем для оборони громадян та об’єктів критичної інфраструктури.

Зв'язок роботи з науковими програмами, планами, темами. Частина дисертаційної роботи виконано в рамках договору про співпрацю між кафедрою Прикладної радіоелектроніки Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» та підприємством ТОВ «ТІНСТРУМ» за договором № Дндч/0201.01/2100.02/47/2023.

Мета й завдання дослідження.

Метою дисертаційної роботи є підвищення ефективності системи виявлення та локалізації малих безпілотних літальних апаратів та обробки їхніх акустичних сигналів зареєстрованих вхідним сенсором.

Основними ***завданнями*** дослідження є:

1) Проведення аналізу основних методів виявлення та локалізації малих безпілотних літальних апаратів;

2) Створення та накопичення бази аудіозаписів акустичних сигнатур різних моделей малих безпілотних літальних апаратів;

3) Розроблення математичної моделі для ідентифікації типу малих безпілотних літальних апаратів, використовуючи базу аудіозаписів їхніх акустичних сигнатур

4) Розроблення методу цифрової обробки акустичних сигналів із використанням згорткової нейронної мережі глибокого навчання з оптимізатором Adam;

5) На основі запропонованого методу, розробити програмне забезпечення для обробки акустичних сигналів на предмет наявності сигнатур малих безпілотних літальних апаратів;

6) Створення акустичного детектора МЛА.

7) Перевірка ефективності розробленого програмного та апаратного забезпечення для виявлення МЛА.

Об`єктом дослідження є процес акустичного виявлення малих безпілотних літальних апаратів.

Предметом дослідження є методи реалізації системи передачі та оброблення акустичних сигналів малих безпілотних літальних апаратів.

Методи дослідження. У дисертаційній роботі використовуються методи математичної статистики для оцінювання працездатності акустичного детектора, схемотехніки для обґрунтування вибору його елементної бази, спектрального аналізу для дослідження аудіосигналів МЛА, а також машинного навчання для розроблення алгоритму класифікації даних.

Для моделювання архітектури нейронної мережі та обробки результатів дослідження використовувався програмний пакет Matlab.

Наукова новизна отриманих результатів полягає в тому, що:

1) Вперше запропоновано математичну модель ідентифікатора МЛА, наукова новизна якої полягає в тому, що вона ґрунтується на використанні бази аудіозаписів акустичних сигнатур дронів та надає можливість зменшити похибку під час класифікації їхньої моделі.

2) Вперше розроблено метод цифрової обробки акустичних сигналів дронів, наукова новизна якого полягає в тому, що він ґрунтується на теорії штучного інтелекту та використовує згорткову нейронну мережу глибокого навчання з оптимізатором Adam для підвищення швидкості та ефективності ідентифікації типу малих безпілотних літальних апаратів.

3) Удосконалено методику розрахунку кількості мел-фільтрів та їхнього частотного діапазону, що, на відміну від існуючих, враховує бажану частотну розділову здатність та дає змогу підвищити швидкість та точність отримання

основних представлень аудіосигналу МЛА під час використання техніки зважування коефіцієнтів ШПФ за допомогою банків мел-фільтрів.

Практичне значення отриманих результатів полягає в тому, що:

1) На основі розробленої математичної моделі, створено базу аудіосигналів малих безпілотних літальних апаратів моделей «*Mavic 2 Pro*», «*Mavic 3*» та «*FPV*», що використовуються військовими в розвідувальних операціях, а також цивільного квадрокоптера «*Feilun Fx137*».

2) На основі розробленого методу та методики, створено програмне забезпечення для швидкого та ефективного оброблення аудіосигналів малих безпілотних літальних апаратів із використанням згорткової нейронної мережі глибокого навчання з оптимізатором Adam.

3) Створено акустичний детектор малих безпілотних літальних апаратів з ефективною відстанню виявлення об’єктів до 200 метрів.

4) Надано відповідні універсальні рекомендації для подальшого вдосконалення системи детекції малих безпілотних літальних апаратів.

Практичний результат дисертаційного дослідження може бути використаний під час побудови ефективних антидронних систем для оборони громадян та об’єктів критичної інфраструктури.

Частину практичних результатів дисертаційної роботи отримано та впроваджено на підприємстві ТОВ «ТІНСТРУМ» (м. Київ) за договором № Дндч/0201.01/2100.02/47/2023.

Особистий внесок. Основні теоретичні розрахунки, розроблення програмного забезпечення для обробки аудіосигналів, моделювання архітектури нейронної мережі та перевірка ефективності роботи прототипу акустичного детектора МЛА, виконані автором самостійно.

Мета, завдання наукового дослідження та висновки було сформульовані в співавторстві з науковим керівником.

У науковій статті *«Огляд методів виявлення та локалізації малих безпілотних літальних апаратів»* здобувачем особисто проведено аналіз основних методів виявлення та локалізації МЛА, а також теоретичні розрахунки

максимальної дальності виявлення дрону моделі «*DJI Mini 2 Fly More Combo*» кожним із методів. У роботі *«Електроакустичний тракт детектора для виявлення малих безпілотних літальних апаратів»* разом із науковим керівником було створено прототип акустичного детектора та проаналізовано його елементу базу. Самостійно автором було проведено експериментальні дослідження по створенню та наповненню бази акустичних шумів МЛА. У статті *«Алгоритм оброблення аудіосигналів із використанням методу машинного навчання»* автором самостійно було розроблено алгоритм аналізу та класифікації акустичних сигналів із використанням згорткової нейронної мережі, а також проведено тестування його роботи з використанням моделі МЛА «*DJI Mavic 3*».

Апробація матеріалів дисертації відбулася на 3-х міжнародних конференціях:

1. III Міжнародній науково-практичній конференції «Філософія і наукова-технічна творчість у хронотопі технічного університету», Київ, Україна, 2020.

2. X Міжнародній науково-технічній конференції «Радіотехнічні поля, сигнали, апарати та системи – 2021» (РТПСАС-2021), Київ, Україна, 2021

3. XI Міжнародній науково-технічній конференції «Радіотехнічні поля, сигнали, апарати та системи – 2022» (РТПСАС-2022), Київ, Україна, 2022.

Публікації. За результатами дисертаційної дослідження опубліковано 6 наукових робіт: 3 статті в фахових виданнях України, з яких 1 у періодичних виданнях, що включенні до міжнародної наукометричної бази WEB OF SCIENCE, та 3 тези доповіді в збірниках матеріалів міжнародних науково-технічних конференцій.

Структура й обсяг дисертації. Дисертація складається зі вступу, чотирьох розділів у основній частині, загальних висновків, списку використаних джерел та трьох додатків. Загальний об’єм дисертаційної роботи складає 183 сторінки, з яких 152 сторінки основного тексту, 70 рисунків та 9 таблиць. Список використаної в роботі літератури включає в себе 50 джерел. Додатки розміщено на 31 сторінці.

РОЗДІЛ 1. АНАЛІЗ ОСНОВНИХ МЕТОДІВ ВИЯВЛЕННЯ ТА ЛОКАЛІЗАЦІЇ МАЛИХ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ

1.1. Огляд завдання виявлення та локалізації малих безпілотних літальних апаратів

Згідно із визначенням, що схвалене Міжнародною організацією цивільної авіації, безпілотний літальний апарат («БПЛА» або «дрон») представляє собою повітряне судно без пілота, що дистанційно керується із землі, з борту іншого судна, або запрограмоване та повністю автономне (рис. 1.1).

Першою появою військових бойових безпілотників учені вважають 12 липня 1849 року, коли під час Італійської революції австрійські війська провели бомбардування бунтівної Венеціанської республіки з аеростатів, які були оснащені механізмами автоматичного скидання боєприпасів.

Рисунок 1.1 — БПЛА *«DJI Mavic 3»*

Хоча ідея появи безпілотних літальних апаратів виникла досить давно, але в сьогоднішньому вигляді вони з’явилися тільки в останні десятиліття, у період активного розвитку робототехніки та штучного інтелекту. Це стало можливим завдяки розвитку та вдосконаленню низки технологій насамперед супутникової радіонавігації, радіолокації, оптичних, акустичних та інфрачервоних сенсорів, а також появі на борту БПЛА досить потужних обчислювальних засобів, які дають змогу інтерпретувати дані, отримані із цих видів сенсорів.

Зараз є багато різновидів дронів, тому базуючись на їхніх основних характеристиках Міжнародною асоціацією з питань безпілотних літальних систем AUVSI (*Association for Unmanned Vehicle Systems International*) була запропонована універсальна класифікація БПЛА (*табл. 1.1*) [2].

Таблиця 1.1 — «Універсальна класифікація БПЛА за AUVSI»

Група	*Категорія*	*Злітна маса, кг*	*Дальність польоту, км*	*Висота польоту, м*	*Тривалість польоту, г*
Малі БПЛА	Nano БПЛА	< 0,025	< 1	100	< 0,5
	Micro БПЛА	< 5	< 10	250	1
	Mini БПЛА	20 – 150	< 30	150 – 300	< 2
	Легкі БПЛА для контролю переднього краю оборони	25 – 150	10 – 30	3000	2 – 4
	Легкі БПЛА з малою дальністю польоту	50 – 250	30 – 70	3000	3 – 6
	Середні БПЛА	150– 500	70 – 200	5000	6 – 10
Тактичні	Середні БПЛА з великою тривалістю польоту	500 – 1500	>500	8000	10 – 18
	Маловисотні БПЛА для проникнення в глибину оборони противника	250 – 2500	>250	50 – 9000	0,5 – 1
	Маловисотні БПЛА з великою тривалістю польоту	15 – 25	>500	3000	>24
	Середні БПЛА з великою тривалістю польоту	500 – 1500	>500	8000	10 – 18
	Маловисотні БПЛА для проникнення в глибину оборони противника	250 – 2500	>250	50 – 9000	0,5 – 1
	Маловисотні БПЛА з великою тривалістю польоту	15 – 25	>500	3000	>24
	Середньовисотні БПЛА з великою тривалістю польоту	1000 – 5000	> 500	5000 – 8000	24 – 48
Тактичні	Висотні БПЛА з великою тривалістю польоту	2500 – 5000	> 2000	20000	24 – 48
Стратегічні	Бойові (ударні) БПЛА	>1000	1500	12000	2
	БПЛА, оснащені бойовою частиною (літального дії)	.	300	4000	3 – 4
	БПЛА-помилкові цілі	150– 500	0 – 500	50 – 5000	< 4
Спец. признач.	Стратосферні БПЛА	> 2500	> 2000	> 20000	> 48
	Екзостратосферні БПЛА	.	.	> 30500	.

Наведена вище класифікація на сьогодні поширюється, як на вже наявні моделі дронів, так і на ті, що будуть розроблятись у майбутньому.

Малі безпілотні літальні апарати через свої габаритні розміри здатні виконувати широкий спектр корисних для людства функцій: передусім це моніторинг та контроль територій, аеро фото або відеозйомка, крім того, дрони можуть здійснювати доставку невеликих вантажів, а також брати участь у проведенні пошуково-рятувальних операціях.

Незважаючи на це, усе частіше фіксуються правопорушення під час використання МЛА, наприклад, доставка заборонених предметів, бойових снарядів, контрабанди, напади на людей, несанкціонована зйомка державних воєнних об’єктів та заборонених територій, що призводять до витоку конфіденційної інформації. У світлі цього важливим і актуальним стає *науково-технічне завдання* — аналіз основних методів та засобів для виявлення малих безпілотних літальних апаратів, розроблення детектора з одним вхідним сенсором-мікрофоном, що дасть змогу моніторити контрольовану зону та своєчасно розгортати контрзаходи, усуваючи потенційні неприємні наслідки.

1.2. Методи виявлення і локалізації МЛА

Є три основні технічні методи детектування малих безпілотних літальних апаратів — оптичний, радіолокаційний та акустичний.

1.2.1. Оптичний метод

Перспективним та досить інформативним підходом до виявлення МЛА є використання систем оптичної локації [3], які, як правило, містять пасивні сенсори різних оптичних діапазонів. Джерелами електромагнітного випромінювання оптичного діапазону для цих пристроїв є лазери, що працюють в імпульсному або в безперервному режимі. Деякі оптичні пристрої служать, як для формування передавання випромінювання лазера, так і для приймання відбитого від об'єкта спостереження лазерного випромінювання.

Інколи в системах оптичної локації застосовуються індивідуально адаптовані оптичні формуючі пристрої для пасивних сенсорів, що працюють у різних діапазонах довжин хвиль, та скануючий пристрій (камера спостереження

високої роздільової здатності), що забезпечує огляд заданого простору. Для функціонування системи застосовуються обчислювальні пристрої, що забезпечують обробку отриманих вхідних сигналів, проведення вбудованого контролю даних і видавання необхідної інформації в зовнішній пристрій.

Можна виділити такі недоліки оптичних систем:

- невеликі кути огляду простору за азимутом;
- малі швидкості сканування заданого простору контрольованої зони;
- недостатній захист сенсорів від зовнішніх і внутрішніх засвічень.

Оптичні системи зазвичай працюють у двох режимах сканування — пасивному та активному.

У *режимі пасивного сканування* камера з ІЧ сенсором (лінійка фотоелементів) повертається на відповідний кут місця, скануючи навколишній простір. Під час виявлення об'єкта з температурним контрастом сигнал з ІЧ сенсора подається на процесор (ПР), який керує взаємодією всіх блоків і пристроїв за зовнішніми командам. Команда з ПР подається на телевізійний канал нічного бачення (ТКНБ) для спрацьовування електронного затвора електрооптичного перетворювача світла (ЕОП). Отримана інформація з ТКНБ надходить у ПР, який визначає величину кутової неузгодженості між «істинним» об'єктом і його температурним аналогом. Кутові координати «істинного» об'єкта передаються в зовнішній пристрій. За відсутності інформації з ТКНБ про «істинне» положення об'єкта, що спостерігається, у зовнішній пристрій видаються кутові координати об'єкта з температурним контрастом.

Режим активного сканування включає в себе підсвічування атмосфери лазерним випромінюванням, приймання відбитого променя і використання обчислювально-аналітичної системи. Сканування атмосфери в тривимірному просторі здійснюють за допомогою керованого скануючого лазера. МЛА або турбулентні потоки, спричинені ними, відхиляють у просторі промінь когерентного джерела світла.

Відхилення променю лазера контролюють телеметричним пристроєм. Технічний результат — підвищення ймовірності виявлення об'єктів і підвищення

точності вимірювання просторових координат цілі. Прикладами таких систем є наземні лазерні далекоміри, наземні локатори, лазерні системи розвідки тощо.

Зазвичай під час застосування системи оптичної локації з камерами спостереження високої розділової здатності формуються зображення об’єктів [3]. Далі відбувається обробка таких зображень і виділяються ділянки відповідні малому безпілотному літальному апарату.

Важливо зауважити, що обробку зображень необхідно здійснювати в реальному масштабі часу. Також значна проблема з’являється за умови моніторингу величезної території контрольованої зони: на досить великій відстані МЛА мають малі кутові розміри й на світлочутливій матриці камери вони будуть займати невелику кількість пікселів, тому їх легко переплутати зі звичайними птахами та подати сигнал хибної тривоги (рис. 1.2).

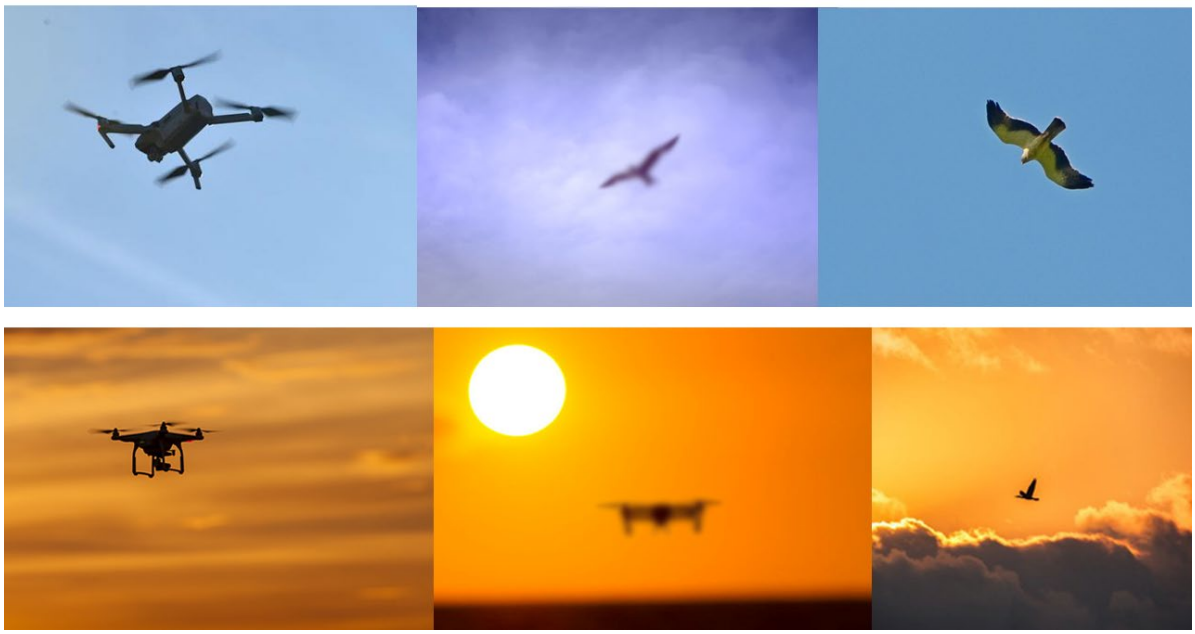


Рисунок 1.2 — Зображення МЛА та птахів за різної розділової здатності

Саме тому і використовують спеціальні алгоритми цифрової обробки зображень для виділення малих безпілотних літальних апаратів на тлі досить складних фонових об'єктів. Приклад алгоритму вказано на рисунку 1.3.

Оптичне детектування дронів значно залежне від чинників оточуючого середовища. Очевидно, що погіршення результатів виявлення МЛА відбувається через збільшення зони огляду та поля зору, обмеження прямої видимості, наприклад, через сильний туман, пил або з настанням ночі.

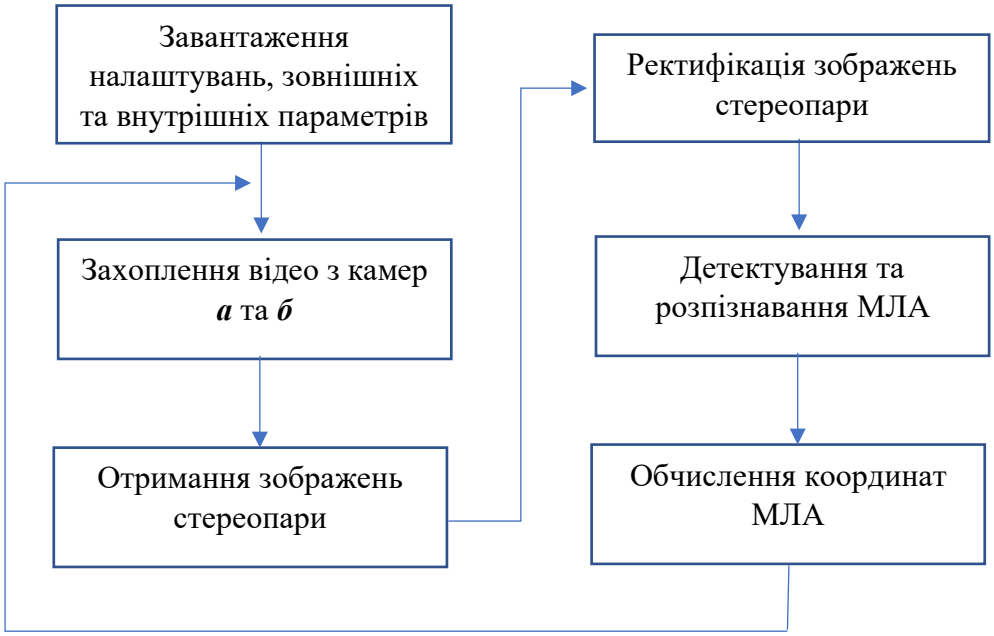


Рисунок 1.3 — Алгоритм обробки зображень

Дальність детектування об’єктів завжди менша або рівна деякому значенню R_{max}, що відповідає найбільш сприятливому поєднанню фізичних умов спостереження. За такої умови закон розподілу дальності детектування найбільш повно характеризує засоби й умови спостереження [4]. Дальність виявлення МЛА оптичними системами визначають за таким виразом [5]:

$$R_{\text{опт}} = \sqrt{\frac{B_\lambda \cdot S_{\text{ц}} \cdot k_{\text{ц}} \cdot S_{\text{пр}} \cdot \Delta\lambda \cdot \tau_{\text{пр}} \cdot \tau_{\text{сер}}}{4\pi P_{\text{пор}_0}}} \quad (1.1)$$

де B_λ – спектральна щільність випромінювання поверхні малих безпілотних літальних апаратів завдяки освітленню її сонцем;

$S_{\text{ц}}$ – ефективна поверхня відбиття цілі в оптичному діапазоні;

$k_{\text{ц}}$ – коефіцієнт відбиття поверхні дрону;

$S_{\text{пр}}$ – ефективна площа об’єктива, на яку падає випромінювання;

$\Delta\lambda$ – смуга пропускання оптичного фільтра (мкм);

$P_{\text{пор}_0}$ – порогова чутливість приймального пристрою;

$\tau_{\text{пр}}, \tau_{\text{сер}}$ – коефіцієнти пропускання приймального пристрою і середовища відповідно.

Розрахуємо максимальну дальність виявлення МЛА оптичною системою на прикладі квадрокоптера *«DJI Mavic 3»* (рис.1.1). Його розміри у «відкритому» стані (з пропелерами) : *347×283×107 мм* (Д×Ш×В).

Будемо вважати, що процес детектування відбувається за нормальних погодних умов, у ясний сонячний день, тому спектральна щільність випромінювання поверхні цілі $B_\lambda = 10^2$ *Вт/ см*^{2} · *мкм* · *ср.*

Визначимо значення ефективної поверхні відбиття (ЕПВ) дрону. Відбивальні властивості поверхні об'єктів зазвичай можна охарактеризувати й через ефективну поверхню відбиття (ЕПР) [6]. Щоб виміряти ЕПР із високою точністю, необхідно взяти реальні об’єкти та провести експерименти на полігонах або в радіочастотних камерах без відлуння. За браком необхідного обладнання, ЕПР для нашого дрону розраховано теоретично на прикладі моделі БПЛА *RQ-1 «Predator»*. У [7] авторами проведені необхідні експерименти та зроблено висновок, що для моделі *RQ-1 «Predator»* кругова медіанна ЕПР (яке використовується під час розрахунків дальності виявлення БПЛА з ймовірністю 0,5) складає *0,93 м*^{2}. Наша модель квадракоптеру *«DJI Mavic 3»* менше наведеної *RQ-1 «Predator»* у 24 раз, відповідно значення ЕПР = *0,039 м*^{2}.

Коефіцієнт відбиття поверхні $k_ц = 0.9$, вважаємо, що він не залежить від кута падіння світла на об’єкт, а також, що корпус малого безпілотного літального апарату буде максимально контрастного кольору до оточуючого середовища.

Діаметр об’єктиву камери виявлення $d = 50$ *мм* та має форму круга, тому за ідеальних умов, ефективна площа, на яку падає випромінювання дорівнює: $S_{\text{пр}_0} = \pi d^2/4 = 1{,}963\ \text{мм}^2$.

У якості смуги пропускання оптичного фільтра будем використовувати видимий діапазон довжин хвиль: границя коротких хвиль 380 … 400 нм, границя довгих хвиль 760…780 нм. Тому $\Delta\lambda$ = 780 – 380 = *500 нм.*

Приймальний пристрій оптичної системи буде використовувати аналогову відеокамеру високої чутливості *DN-9770D*, що має 1/3-дюймову ПЗС-матрицю Sony 960H EXview HAD CCD II та процесор для оброблення відео Sony DSP

Effio-E (рис. 1.4). Основні технічні характеристики відеокамери неведені в таблиці 1.2 [8].

Камера має динамічний діапазон 52 дБ і високе розширення 680/700 ТВЛ у кольоровому / чорно-білому режимах. Програмний режим «день / ніч» дає змогу відеокамері формувати кольорове / чорно-біле зображення на мінімальних рівнях освітленості (0.03 / 0.01 лк), а система придушення шуму 2D-NR забезпечує чітке зображення в разі зниженого освітлення.

Рисунок 1.1 — Вулична камера спостереження DN-9770D [8]

Відповідно порогова чутливість камери $P_{\text{пор}_0}$ під час формування кольорового зображення дорівнюватиме *0.03 лк* (при співвідношенні сигнал/шум (С/Ш) 52 дБ) або $4.39 \cdot 10^{-9}$ *Вт / см*^{2}.

Таблиця 1.2 — «Технічні характеристики відеокамери *DN-9770D*» [8]

Найменування параметра	***Значення***
1	**2**
Сенсор зображення	«День/ніч» (Easy day/night) з WDR
Чутливий елемент	1/3” ПЗС SONY 960H EXview HAD CCD II
Кількість пікселів матриці	976 x 582
Розширення камери	Кольорове: 680 ТВЛ; Ч/б: 700 ТВЛ
Мінімальна освітленість	Кольорове: 0.03 лк (F1.6); Ч/б: 0.01 лк (F1.6)
Процесор	DSP Sony Effio-E
Співвідношення сигнал/шум	Більше 52 дБ (АРП вимк.)
Гамма-корекція	0,45
Електронний затвор відеокамери	Авто 1/50-1/100000 с, Ручний 1/50, FL 1/120, 1/250, 1/500, 1/1000, 1/2000, 1/4000, 1/10000 с.
Режим «день-ніч» (програмний)	Авто (регульований поріг перемикання) / Колір/Ч/б сумісність з ІЧ-прожектором
Дальність ІЧ підсвічування	Не менше 15 м
Кут ІЧ підсвічування	Комбінований, не менше 80º та 60º (на рівні 1/2 яскравості)
Динамічний діапазон (WDR)	Програмний, 128 x (52 дБ)
Відеовихід	Композитний 75 Ом, 1 В; BNC
Компенсація зустрічного засвічення	HLC/BLC/OFF
Шумоподавлення	2D-NR
Тип об'єктива	Зі змінним фокусом
Фокусна відстань об'єктива	f=2,8 - 12 мм; F=1.6

Продовження таблиці 1.2.

1	2
Кріплення об'єктиву	C/CS (з адаптером)
Синхронізація	Внутрішня
Живлення відеокамери	10,8 - 13,2 В (пост. струм) / 24 В (змін. струм)
Діапазон робочих температур	Від -30 °C до +50 °C
Вологість (макс.)	80% (без конденсату)

Коефіцієнти пропускання приймального пристрою і середовища приймемо рівними 1, вважаємо, що ефект затухання не відбувається та не має розсіювання.

Підставимо знайдені параметри у формулу (1.1):

$$R_{\text{опт}} = \sqrt{\frac{10^2 \cdot 0{,}39 \cdot 0{,}9 \cdot 1{,}963 \cdot 10^{-3} \cdot 500 \cdot 10^{-3} \cdot 1 \cdot 1}{4\pi \cdot 4{,}39 \cdot 10^{-9}}} = 250 \text{ м}$$

Оптичний метод виявлення малих безпілотних літальних апаратів значно залежний від чинників оточуючого середовища. Він використовується у світлу пору доби для підвищення розділової здатності камер відеоспостереження.

<u>Перевагами</u> цього способу детекції дронів є:

- *Виявлення МЛА, що не мають радіочастотної передачі;*
- *Оптичні датчики є пасивними пристроями;*
- *Велика дальність розпізнавання цілі;*
- *Низька вартість оптичних сенсорів.*

На якість та максимальну відстань детектування безпілотних літальних апаратів оптичними сенсорами впивають:

- *Розміри, форми та матеріал корпусу МЛА;*

Зі зменшенням розміру безпілотних літальних апаратів, зростає складність їхнього виявлення та зменшується максимальна відстань успішної детекції об’єктів. Форма дронів та матеріал поверхні корпусу впливає на його освітленість та контрастність до оточуючого середовища.

- *Впливи атмосфери;*

Забруднення атмосфери суттєво погіршує видимість для оптичних сенсорів, але через водяні пари в атмосфері сонячні промені відбиваються вгору, що зможе підвищити рівень освітленості МЛА.

- *Зона огляду;*

Максимальна відстань детектування цілі досягається через звуження поля зору камер спостереження, це зменшує зону огляду та збільшує час пошуку.

- *Погодні умови та доба дня;*

Малі безпілотні літальні апарати неможливо виявити в хмарну погоду, туман та якщо освітленість фону оточуючого середовища й об'єкта однакові. Також оптичні системи не ефективні після заходу сонця.

Для збільшення імовірності детектування дронів у нічний час, зазвичай використовують тепловізійні камери. У цьому випадку зображення формуються в інфрачервоному діапазоні електромагнітних хвиль. Основне тепло виділяється зазвичай силовою установкою МЛА, у меншому ступені електронними компонентами, а також точками гальмування на несучих краях крил, пропелерів і гвинтів [3]. Розробники безпілотників, щоб запобігти випроміненню в ІЧ діапазоні в напрямку розташованих на землі приймачів, намагаються або направити його в сторону неба, створюючи нові спеціальні конструкції дронів, або використовувати матеріали з низькою випромінювальною здатністю, наприклад, срібло та алюміній [9]. Відповідно МЛА можливо виявити за їхньою випромінювальною здатністю, контрастом та площею випромінювання.

Розподіл інтенсивності випромінювання для реальних (сірих) тіл $J_{\text{с.т.}}(\lambda, T)$ визначають за законом Планка [10]:

$$J_{\text{с.т.}}(\lambda, T) = \xi \cdot \frac{2\pi hc^2}{\lambda^5} \cdot \frac{1}{e^{\left(\frac{hc}{kT\lambda}\right)} - 1} \tag{1.2}$$

де $h = 6{,}62 \cdot 10^{-34}$ Дж·с — стала Планка;

$c = 3 \cdot 10^8$ м/с — швидкість поширення електромагнітних хвиль у вакуумі;

λ – довжина хвилі;

$k = 1{,}38 \cdot 10^{-23}$ Дж/К — стала Больцмана;

T – абсолютна температура нагрітого тіла;

ξ – коефіцієнт випромінювання.

Атмосфера поглинає інфрачервоне випромінювання краще, ніж світло. Але є так звані атмосферні вікна для ІЧ діапазону з межами довжин хвиль 3 – 4

мм та 8 – 12 мм, тому інфрачервоні пристрої проектують так, щоби приймати в одному з «вікон», не дивлячись на те, у якому з них розташовується максимум випромінювання МЛА згідно із законом Планка.

Обробка одержуваного зображення відбувається за допомогою спеціальних алгоритмів, виділяються ділянки, що відповідають малим безпілотним літальним апаратам, тобто відбувається їхнє виявлення.

У низькі випадків для того, щоби підвищити ефективність оптичного інфрачервоного методу використовують «комплексування», тобто на основі оптичного інфрачервоного зображення формується загальне інтегральне зображення з огляду на дані з парціальних зображень.

Максимальна дальність виявлення малих безпілотних літальних апаратів (з температурою *T*) пасивними інфрачервоними системами, визначають за [5] :

$$R_{\text{іч}} = \sqrt{\frac{S_{\text{ціч}} \cdot S_{\text{пріч}} \cdot \tau_{\text{пріч}}}{P_{\text{поріч}}} \int_{\lambda_1}^{\lambda_2} \tau_{\text{сріч}}(\lambda) \cdot J_{\text{с.т.}}(\lambda) d\lambda} \qquad (1.3)$$

де $S_{\text{ціч}}$ – площа проєкції МЛА на картинну площину, перпендикулярну напрямку спостереження цілі;

$S_{\text{пріч}}$ – ефективна площа приймальної апертури ІЧ пристрою;

$\tau_{\text{пріч}}$ – коефіцієнти пропускання приймального пристрою;

$P_{\text{поріч}}$ – порогова чутливість приймального пристрою;

λ_1 , λ_2 – межі спектрального діапазону приймального пристрою;

$\tau_{\text{сріч}}(\lambda)$ – спектральний коефіцієнт пропускання середовища.

За формулою (1.3) розрахуємо максимальну дальність детектування квадрокоптера *«DJI Mavic 3»* за допомогою ІЧ системи. За законом Планка необхідно визначити розподіл інтенсивності випромінювання $J_{\text{с.т.}}(\lambda, T)$ для сірого тіла (1.2) на довжині хвилі $\lambda = 2$ мкм. Матеріал корпусу дрона — пластик, тому згідно з [11] коефіцієнт випромінювання $\xi = 0.95$, за температури оточуючого середовища 20°С абсолютна температура нагрітої ділянки малого безпілотного літального апарату дорівнюватиме 298 К, відповідно значення інтенсивності випромінювання дрону:

$$J_{\text{с.т.}} = 0{,}95 \cdot \frac{2\pi \cdot 6{,}62 \cdot 10^{-34} \cdot (3 \cdot 10^{8})^{2}}{(2 \cdot 10^{-6})^{5} \cdot e^{\left(\frac{6{,}62 \cdot 10^{-34} \cdot 3 \cdot 10^{8}}{1{,}38 \cdot 10^{-23} \cdot 298 \cdot 2 \cdot 10^{-6}}\right)} - 1} = 362{,}384\ \frac{\text{Дж}}{\text{с} \cdot \text{м}^{3}}$$

За формулою (1.3) максимальна дальність виявлення малого безпілотного літального апарату *«DJI Mavic 3»* дорівнює:

$$R_{\text{іч}} = \sqrt{\frac{0{,}039 \cdot 1{,}963 \cdot 10^{-3}}{4{,}39 \cdot 10^{-9}} \int_{0{,}78 \cdot 10^{-6}}^{1000 \cdot 10^{-6}} 1 \cdot 362{,}384\, d\lambda} = 79{,}5\ \text{м}$$

Інфрачервоний метод виявлення має схожі переваги з оптичним, але ІЧ спосіб детектування дронів використовується вночі.

З <u>недоліків</u> цього методу можна виділити такі:

- Спеціальні *конструкції* та *матеріал корпусу* для зменшення випромінювальної здатності, контрастності та площі випромінювання малих безпілотних літальних апаратів*;*
- *Невисока максимальна дальність виявлення цілі;*
- *Інфрачервоне випромінювання* добре *поглинається атмосферою,* окрім вікон із межами довжин хвиль 3 – 4 і 8 –12 мм.
- *Вузька зона огляду.*

Бездоганні інфрачервоні пристрої спостереження перебувають, лише на стадії розроблення, а детектування МЛА зі слабким випромінюванням можливе, лише за малих тілесних кутів, приблизно 5°. На практиці ефективний процес детектування дронів потребує створення спеціальних високочутливих ІЧ пристроїв, розроблення спеціальних алгоритмів огляду простору та прийняттю рішення про наявність цілі або її відсутність у зоні контролю. Основними технічними характеристиками таких систем є дальність її дії та час огляду. Ці характеристики визначають за ступенем оптимізації всіх елементів і алгоритмів системи детектування безпілотних літальних апаратів. Крім того, для контролю великих тілесних кутів потрібні багатопроменеві інфрачервоні системи, у яких деяка кількість нерухомих променів "очікують" прольоту цілі [12].

1.2.2. Радіолокаційний метод

Радіолокаційний метод використовується у двох режимах — *активної* і *пасивної* радіолокації.

Активний метод радіолокації є досить ефективним, оскільки має досить великий імпульсний об’єм пошуку, а також значну дальність виявлення [13]. Під час використанні цього методу радіолокаційні станції випромінюють зондуючий сигнал, коли він відбивається від МЛА, то надходить на приймач станції. Далі відбувається процес аналізу сигналу, звідки ми можемо визначати просторові координати, дальність знаходження малого безпілотного літального апарату, а також одержувати деяку додаткову інформацію про об'єкт.

Зокрема, у сигналі міститься інформація про сам об'єкт, його тип, кількість гвинтів, ступень завантаженості. Така інформація знаходиться в мікро-доплерівській сигнатурі, яка «витягається» із сигналу, що надходить на вхід системи (рис. 1.5).

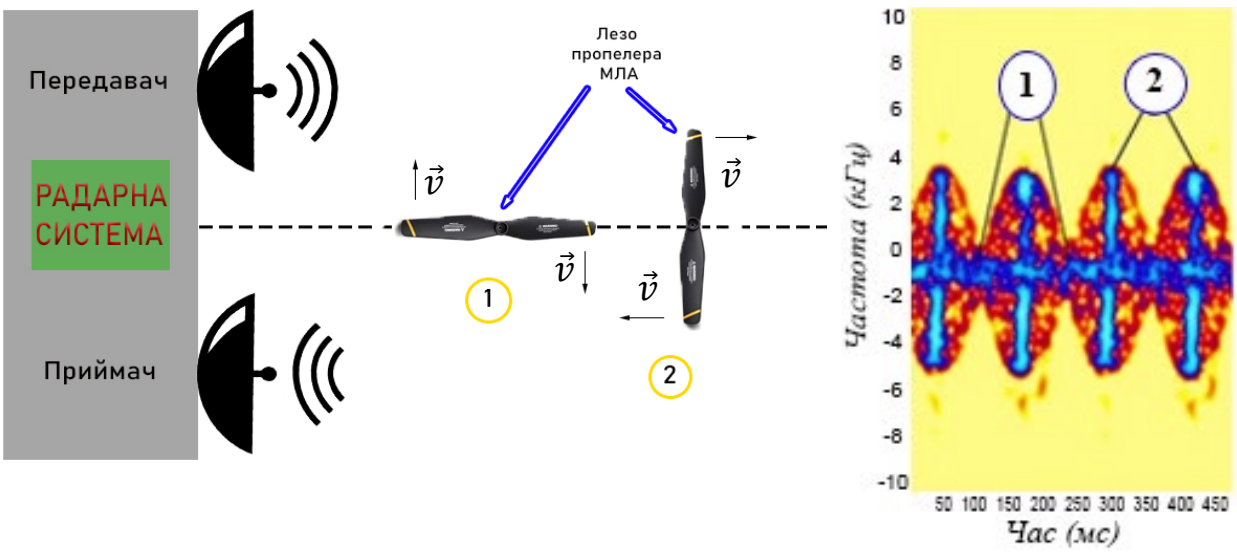


Рисунок 1.2 — Мікро-доплерівська модуляція сигналу МЛА

Радіолокаційні системи дають змогу детектувати дрони на досить великій відстані, приблизно десятки кілометрів, якщо вони досить великі за розміром (більша площа відбиття сигналу) та мають значну вагу, але максимальна дальність виявлення значно зменшується в міру зменшення ваги та розмірів

безпілотних літальних апаратів. Для МЛА вона приблизно становить одиниці кілометрів або сотні метрів. У таких умовах буде недостатньо часу, щоби призвести методи протидії у функціональну готовність. Також на якість детектування впливає і те, що багато безпілотних літальних апаратів виготовляють із композитних матеріалів, тому електромагнітні хвилі проходять крізь поверхню дронів та, лише частково, відбиваються від неї [14].

Дальність виявлення квадрокоптера «*DJI Mavic 3»* активною РЛС можна визначити за таким виразом [13] :

$$R_{\text{РЛС}_{\text{акт}}} = \sqrt[4]{\frac{P_{\text{пер}} \cdot G_{\text{пер}} \cdot G_{\text{пр}} \cdot \lambda^2 \cdot \sigma_{\text{МЛА}}}{(4\pi)^3 \cdot P_{\text{пор}_{\text{РЛС}}}}} \tag{1.4}$$

де $P_{пер}$ – потужність передавача;

$G_{пер}$ – коефіцієнт підсилення передавальної антени;

$G_{пр}$ – коефіцієнт підсилення приймальної антени;

$\sigma_{МЛА}$ – ЕПР МЛА;

$P_{\text{пор}_{\text{РЛС}}}$ – порогова потужність сигналу на вході приймача РЛС.

Для розрахунку максимальної дальності виявлення МЛА візьмемо РЛС дециметрового діапазону з потужністю передавача *1.5 кВт*, з коефіцієнтами підсилення передавальної (*2500*) та приймальної (*3000*) антени. ЕПР квадракоптеру «*DJI Mavic 3»* дорівнює *0.039 м*^{2}. Також будемо вважати, що мінімальна потужність сигналу на вході приймального пристрою радіолокаційної станції *1.52 ·10* $^{-12}$ *Вт*. Підставимо значення параметрів у (1.4):

$$R_{\text{РЛС}_{\text{акт}}} = \sqrt[4]{\frac{1{,}5 \cdot 10^3 \cdot 2500 \cdot 3000 \cdot 0{,}25^2 \cdot 0{,}039}{(4\pi)^3 \cdot 1{,}52 \cdot 10^{-12}}} = 9765 \text{ м}$$

Досить інформативним є пасивний метод радіолокації. Під час використання цього методу засоби радіотехнічної розвідки (РТР) виявляють малі безпілотні літальні апарати шляхом приймання та аналізу радіосигналів, які випромінюються радіозасобами, наприклад, для встановлення зв'язку з пультом керування, для передачі відео зображення, одержуваного безпосередньо за

допомогою дрона або для отримання навігаційної інформації. Такі радіоелектронні засоби зазвичай розташовані безпосередньо на борту МЛА.

Але пасивна радіолокація дає змогу встановити, лише напрямок руху безпілотних літальних апаратів, а точність детектування зростає зі збільшення часу спостереження. Проте цей метод забезпечує досить велику дальність виявлення, яка визначається за формулою [15] :

$$R_{\text{РЛС}_{\text{пас}}} = \sqrt{\frac{P_{\text{дж}_{\text{РТР}}} \cdot G_{\text{дж}}(\varepsilon, \beta) \cdot G_{\text{пр}}(\varepsilon, \beta) \cdot \lambda^2}{(4\pi)^2 \cdot P_{\text{пор}_{\text{РТР}}}}} \quad (1.5)$$

де $P_{\text{дж}_{\text{РТР}}}$ – потужність сигналу джерела випромінювання;

$G_{\text{дж}}(\varepsilon, \beta)$ – коефіцієнт підсилення антени джерела випромінювання залежно від сферичних координат;

$G_{\text{пр}}(\varepsilon, \beta)$ – коефіцієнт підсилення антени станції РТР залежно від сферичних координат;

$P_{\text{пор}_{\text{РТР}}}$ – порогова потужність сигналу на вході приймача станції РТР.

Однак МЛА можуть здійснювати політ в автономному режимі, тобто в умовах радіомовчання, або із використанням оптико-волоконних ліній зв'язку, у таких випадках переваги пасивного методу зводиться практично до нуля.

Перевагами ***методів*** із використанням ***радіолокаційних систем*** є:

- *Дальність виявлення МЛА в кілометрах;*
- *Сектор огляду 360°;*
- *Активні датчики* — можуть працювати вдень та вночі;
- *Надають детальну інформацію про об'єкт*: дальність польоту, радіальну швидкість, розміри, типи й навіть форми БПЛА, дають змогу локалізувати його за кутом місця / азимуту на великих дистанціях.

Основними недоліками використання систем радіолокації є:

- *Низька мобільність та прихованість РЛС станцій;*
- *Розмір, форма та матеріал корпусу зменшують відбивальну здатність МЛА;*

У малих безпілотних літальних апаратів менша площа відбиття електромагнітної хвилі, також їхні корпуси зазвичай виготовлять із композитних матеріалів або покривають радіопоглинальними матеріалами, що значно зменшує рівень відбитого сигналу в напрямку приймача станції.

- *РЛС «сліпі» в ближній контрольваній зоні;*

Проте метод пасивної радіолокації буде неефективним, коли малі безпілотні літальні апарати будуть перебувати в режимі автономного польоту або використовувати оптико-волокно для зв’язку з командним пунктом.

- *Варіація конструкцій станцій за призначенням, діапазоном довжини хвиль, методом дії та характером носія.*

1.2.3. Акустичний метод

Наступний метод виявлення малих безпілотних літальних апаратів полягає в прийомі акустичних хвиль, які випромінюються через обертання несучих гвинтів дрону, його двигунів, а також шумів механічного походження.

Акустичні хвилі — це коливання тиску, що розповсюджуються в повітрі (газах), рідині або у твердому середовищі. Акустичні хвилі поширюються значно повільніше, ніж радіохвилі: у повітрі, наприклад, зі швидкістю приблизно *300 м/с*, у воді — майже *1,5 км/с*, у твердих тілах — *3-6 км/с* [16].

Детальна класифікація розподілу акустичних хвиль за частотою коливань зображена на рисунку 1.6.

Рисунок 1.3 — Спектр звукових коливань:

де 1 — землетруси, блискавки та виявлення ядерних вибухів; 2 — акустичний діапазон; 3 — слух тварин; 4 — ультразвукова чистка; 5 — терапевтична ультразвукова діагностика; 6 — медична ультразвукова діагностика; 7 —

акустична мікроскопія; 8 — інфразвук; 9 — діапазон сприйняття (чутливості) людини; 10 — ультразвук.

Акустичні хвилі природного походження складні, тобто несуть із собою різночастотні коливання. Їхні складові характеризують *амплітудно-частотним спектром* — залежність амплітуди коливань від їх частоти. У звукових шумів та завад дуже широкий безперервний діапазон частот.

Для детектування акустичних хвиль використовують елементи, чутливі до швидких коливань зовнішнього тиску. Найчастіше, це легкі мембрани для перетворення коливань тиску повітря, твердого тіла або якоїсь рідини на механічні коливання, які вже перетворюються на електричні сигнали. Чутливі до звукових коливань, що розповсюджуються в повітрі або в газах, датчики називають ***мікрофонами***; чутливі датчики до акустичних коливань у рідинах — ***гідрофони***; а сенсори акустичних хвиль у твердих тілах — ***стетоскопи***.

За своєю структурою акустичні сигнали МЛА являють собою сукупність гармонічних і широкосмугових складових від *100 Гц* до *10-12 кГц*. Кратні гармоніки спектру слідують із частотами $f_i = f_0 \cdot i$, де f_0 – частота запалювання, а $i = 1,2...$ – номер відповідної гармонічної складової. На високих частотах періодичність спектру акустичного випромінювання пропадає та починається процес випадкового походження. «*Лопоті двигуна формують в середовищі послідовність імпульсів тиску, частотний спектр якої, є комбінацією гармонічних і широкосмугових складових*» (рис. 1.7) [17].

Дальність виявлення акустичного сигналу МЛА визначається за наступною формулою:

$$R_{\text{акуст}} = \sqrt{\frac{I_c}{q^2 \cdot I_з}} \cdot \alpha r_c \qquad (1.6)$$

де $I_с$, $I_з$ – інтенсивність сигналу та завади на вході приймача;

q – співвідношення сигналу до завади;

α – коефіцієнт поглинання звуку в повітрі (дБ/м);

r_c – відстань, для якої визначено акустичний тиск випромінювання.

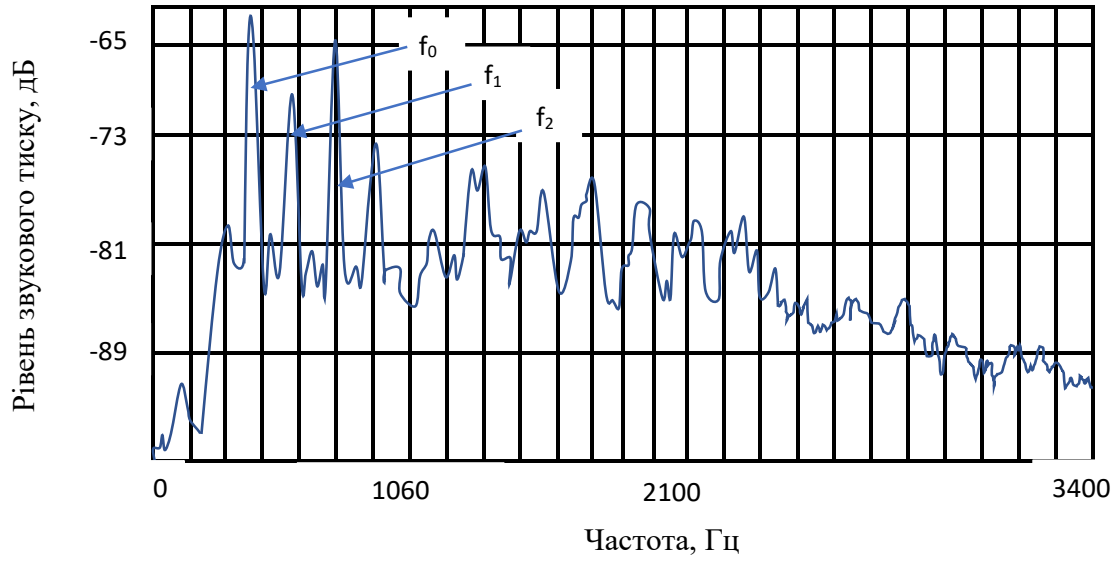


Рисунок 1.4 — Спектр акустичного випромінювання роботи силової установки МЛА *«DJI Mavic 3»*

Проаналізувавши формулу (1.6), можна зробити висновок, що під час зменшення величини вхідного акустичного сигналу, який впевнено реєструє приймач, зростає максимальна відстань детектування джерел випромінювання.

За результатами власного дослідження, проведеного за аналогією до експерименту у [18], за заданої ймовірності виявлення $D = 0.90$ та хибної тривоги $F = 10^{-3}$, співвідношенню С/З за інтенсивностю (параметр виявлення) $q^2 = 0.13$ та різниці рівнів $L = 26$, максимальна відстань виявлення малого літального апарату *«DJI Mavic 3»* приймачем із параметрами налаштування $\Delta f = 2100$ *Гц* (ефективна смуга частот) становить приблизно **78 *м***.

Особливість виявлення малих безпілотних літальних апаратів полягає в тому, що недостатньо провести тільки енергетичне виявлення — виявити деяку матеріальну точку. Також необхідно розв’язати задачу розпізнавання, тобто необхідно відрізнити МЛА від птахів. Локаційні характеристики дронів і птахів дуже близькі, тому це завдання загалом є досить важким.

<u>Перевагами</u> *акустичного методу* є:

- *Пасивність датчиків та їхня низька вартість;*
- *Сектор огляду* контрольованої зони *до 360°* — одночасно можна моніторити всю площу огляду $S_{огл}$;
- *Незалежність від оточуючого рельєфу*;

- *Висока точність виявлення об'єктів.*

З основних недоліків можна виділити такі:

- *Вплив фонових шумів оточуючого середовища;*
- *Малий рівень звукової потужності та широка смуга частот;*
- *Високий рівень апріорної невизначеності відносно структури акустичних сигналів МЛА та завад.*

Проаналізувавши основні методи локалізації, класифікації та відстеження малих безпілотних літальних апаратів можна зробити висновок, що на ефективність детекції дронів значний вплив оказують середовище поширення сигналів, помітність об'єктів у радіолокаційному, інфрачервоному та оптичному діапазонах довжин хвиль та фактори, пов'язані з пристроями спостереження.

Основними показниками ефективності є максимальна дальність детектування цілі R (*рис.1.8*) за заданих ймовірностях правильного виявлення D та хибної тривоги F, розміри площі огляду пристроїв $S_{огл}$, а також площі виявлення МЛА $S_{виявл.}$ [19].

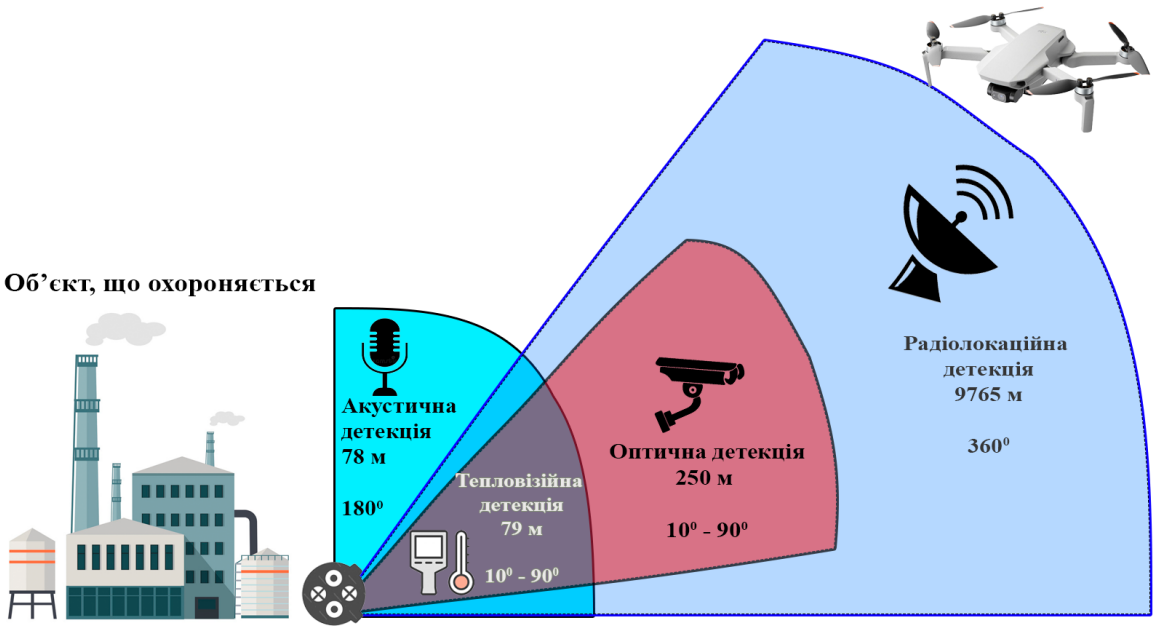


Рисунок 1.5 — Сектор огляду та максимальна дальність детектування МЛА *«DJI Mavic 3»* основними методами

Як видно з рисунку 1.8, методи з використанням систем радіолокації показали найкращу максимальну відстань для успішного виявлення МЛА, навіть якщо він працює в автономному режимі (не випромінює радіочастотні сигнали),

але основним недоліком є те, що всі РЛС «сліпі» в ближній зоні, тому, якщо дрон буде пролітати на близькій відстані до станції, його виявлення буде неможливим.

Через невисокий рівень ефективності пошуку малих безпілотних літальних апаратів, оптичний метод краще використовувати, як допоміжній до більш дієвого методу, наприклад, до радіолокаційного.

Акустичні реєстратори сигналів, через їхню малу максимальну дальність виявлення, але велику площу огляду та можливість працювати в будь-яку добу дня, також можуть бути використані, як додатковий метод детектування і супроводження МЛА, що літають на низькій висоті до поверхні землі.

Підсумовуючи вищесказане, для швидкого та ефективного виявлення малих безпілотних літальних апаратів пропонується використовувати комбіновані системи, які надають дані, одночасно отримані від декількох різних типів датчиків. Радіолокаційні станції необхідні для детектування дронів на великих відстанях до об’єктів, а для захисту «ближньої» контрольованої зони пропонується встановити акустичні датчики.

1.3. Постановка завдання дослідження

З результатів аналізу основних методів виявлення випливає, що малі безпілотні літальні апарати мають невисокий рівень інфрачервоного випромінювання та радіолокаційних сигнатур, у порівнянні, наприклад, з пілотованою авіацією, тому на сьогодні актуальним постає питанням розроблення систем детекції МЛА за допомогою акустичних сигналів із великою максимальною відстанню виявлення до цілі.

Зараз усі типи детекторів дронів використовуються у військовій сфері, що призводить до значного збільшення їхньої вартості через специфіку використання. Саме тому виникають певні проблеми під час купівлі таких пристроїв для охорони, наприклад, цивільних об’єктів. Відповідно створення більш доступного та ефективного аналогу є додатковим підтвердженням актуальності задачі.

Задля зменшення додаткових вимог до пристрою, пропонується процес усієї обробки прийнятого сигналу, включно з класифікацією типу об’єкту, проводити із використанням внутрішніх обчислювальних пристроїв, наприклад, ПК або ноутбуку із відповідними параметрами, для забезпечення необхідної швидкості передачі інформації. Алгоритми детектування та класифікації малих безпілотних літальних апаратів, для збільшення показнику автономності пристрою, мають використовувати, як найменше обчислювального ресурсу. Для збільшення максимальної дистанції виявлення об’єкту необхідне написання програмного алгоритму швидкого та ефективного оброблення акустичних сигналів від МЛА, а саме сигналів із малим співвідношення сигнал/шум.

Метою дисертаційного дослідження є підвищення ефективності системи виявлення малих безпілотних літальних апаратів та обробки їхніх акустичних сигналів зареєстрованих вхідним сенсором.

Додатковою ціллю дисертації є розробка програмного алгоритму для зменшення ймовірності помилки детектування сигнатури дрону під час оброблення акустичних сигналів, а також класифікації цілі. Необхідність створення такого алгоритму з’являється у зв’язку з конфіденційністю інформації про алгоритми, що використовуються в аналогічних системах виявлення дронів, а також за браком матеріалу в наукових статтях та публікаціях.

Таким чином для успішного виконання мети необхідно вирішити наступні ***завдання***:

1. Створення та накопичення бази аудіозаписів акустичних сигнатур різних моделей малих безпілотних літальних апаратів.

2. Розробити математичну модель для ідентифікації типу дрону на фоні шуму оточуючого середовища.

3. Розробити метод обробки акустичних сигналів МЛА на базі використання згорткової нейронної мережі глибокого навчання з оптимізатором Adam.

4. На основі запропонованого методу, розробити програмне забезпечення для обробки акустичних сигналів на предмет наявності сигнатур МЛА.

5. Створити акустичний детектор МЛА.

6. Перевірити ефективність розробленого програмного та апаратного забезпечення для виявлення МЛА.

Висновки до розділу 1

1. Проаналізовано потенційні можливості оптичного, радіолокаційного та акустичного методу детектування МЛА. На основі експериментальних даних для кожного з методів був проведений розрахунок максимальної дальності виявлення малого безпілотного літального апарату моделі *«DJI Mavic 3*».

2. Оптичний метод передбачає використання камер високої розділової здатності у світлий час доби, але він занадто залежний від природних чинників навколишнього середовища. Дальність виявлення малогабаритного дрону моделі *«DJI Mavic 3*» оптичним методом складає біля 250 м.

3. Для виявлення МЛА в темний час доби застосовують ІЧ оптичні пристрої. Основними недоліками підходу є поглинання ІЧ випромінювання атмосферою та мала максимальна дальність детекції (79,5 м для *«DJI Mavic 3*»).

4. Радіолокаційні системи можуть працювати в будь-який час доби та дають змогу виявляти МЛА на відстані приблизно 9,8 км. Основним недоліком є те, що всі РЛС не працюють у ближній зоні.

5. Акустичні сенсори дають змогу ефективно виявляти МЛА, незважаючи на рельєф оточуючого середовища, але результати сильно залежать від наявних акустичних фонових шумів та завад. Ефективна дальність виявлення цілі складає біля 78 м, але, теоретично, може бути підвищена за допомогою спеціальних програмних алгоритмів для оброблення акустичних сигналів.

6. У підсумку аналізу зазначається, що найбільш ефективним підходом до якісного виявлення малих безпілотних літальних апаратів у «ближній зоні» є використання акустичного детектора. Побудова такого типу системи, а також написання комп’ютерного алгоритму для швидкого та якісного оброблення акустичних сигналів і є основною метою та завданням дослідження.

7. Основні положення та результати дослідження за розділом 1 опубліковані в праці [20]

РОЗДІЛ 2. РОЗРОБКА АКУСТИЧНОГО ТРАКТУ ДЕТЕКТОРА МАЛИХ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ

Акустичний детектор МЛА може бути ефективним доповненням до вже наявних систем, що працюють в оптичному чи радіо діапазоні. Детектор складається з електроакустичного тракту, блоку АЦП та блоку цифрового оброблення сигналів на базі ПК (рис. 2.1).

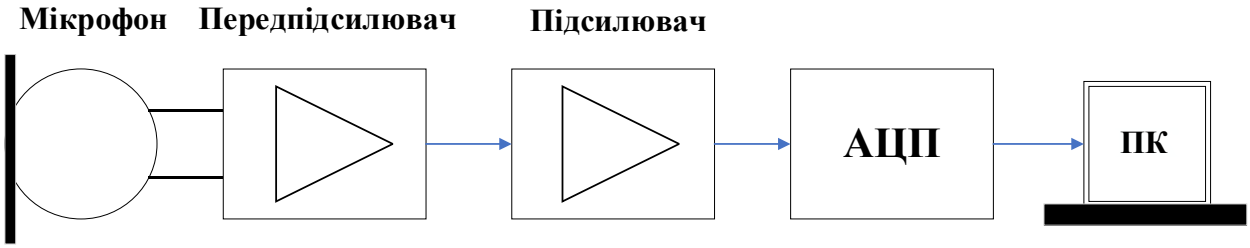


Рисунок 2.1 — Структурна схема акустичного детектора

Одним з найважливіших елементів акустичного реєстратора сигналів є первинний електроакустичний перетворювач, в ролі якого доцільно використовувати мікрофон. Від особливостей конструкції обраного типу мікрофону залежить необхідність використання передпідсилювача. Зареєстровані акустичні сигнатури дронів потрапляють на каскад підсилювача, тому що, залежно від відстані до об’єкту, корисні сигнали цілі потрібно підсилювати або послабляти для подальшої коректної та ефективної обробки на обчислювальному пристрої. Але не можна одразу подавати сигнал на вхід звукової карти ПК, спочатку необхідно перетворити його з аналогового виду в цифровий код, використовуючи АЦП. На ПК уже проходить процес обробки та аналізу акустичного сигналу для виокремлення з нього сигналу МЛА за допомогою спеціальних програмних алгоритмів. Проаналізуємо кожен елемент більш детально для схема-технічної реалізації.

2.1. Вибір акустичного сенсора-мікрофона

Мікрофон — електроакустичний прилад, що перетворює зміну звукового тиску на електричні сигнали. Принцип роботи мікрофона полягає в перетворенні механічних коливань тонкої мембрани в електричну напругу або заряд, завдяки явищу електромагнітної індукції, зміні ємності конденсаторів або п'єзоелектричному ефекту.

Класифікація різних видів мікрофонів показана на рисунку 2.2.

Рисунок 2.2 — Класифікація мікрофонів

Вплив звукового тиску на діафрагму з одної сторони (мікрофон тиску), чи на обидві одночасно, є основною причиною залежності властивостей акустико-

механічних систем. У другому випадку можливий симетричний вплив (мікрофон градієнта тиску) на обидві сторони, або безпосередньо збуджується одна сторона мембрани, а на іншу впливають коливання, які пройшли через системи затримки часу — акустичний або механічний опори (асиметричний мікрофон градієнта тиску). Значний вплив на характеристики мікрофонів оказують механоелектричні частини та шуми.

Вугільні типи мікрофонів функціонують у такий спосіб. Під впливом тиску акустичних сигналів, діафрагма мікрофону починає коливатися та змінювати тиск на зерна вугільного порошку, відповідно змінюється і опір між електродами, що призводить до появи електричного струму в мікрофоні. Такі мікрофони й зараз використовуються в апаратах аналогової телефонії.

У конденсаторному мікрофоні (рис. 2.3) звукові хвилі впливають на тонку мембрану, через що відстань між нею і металевим корпусом змінюється, тим самим, утворюючи конденсатор, що змінює ємність. Якщо подати на пластину постійну напругу, ємність буде змінюватися та викликати струм через конденсатор, тим самим, створюючи електричний сигнал у зовнішньому колі.

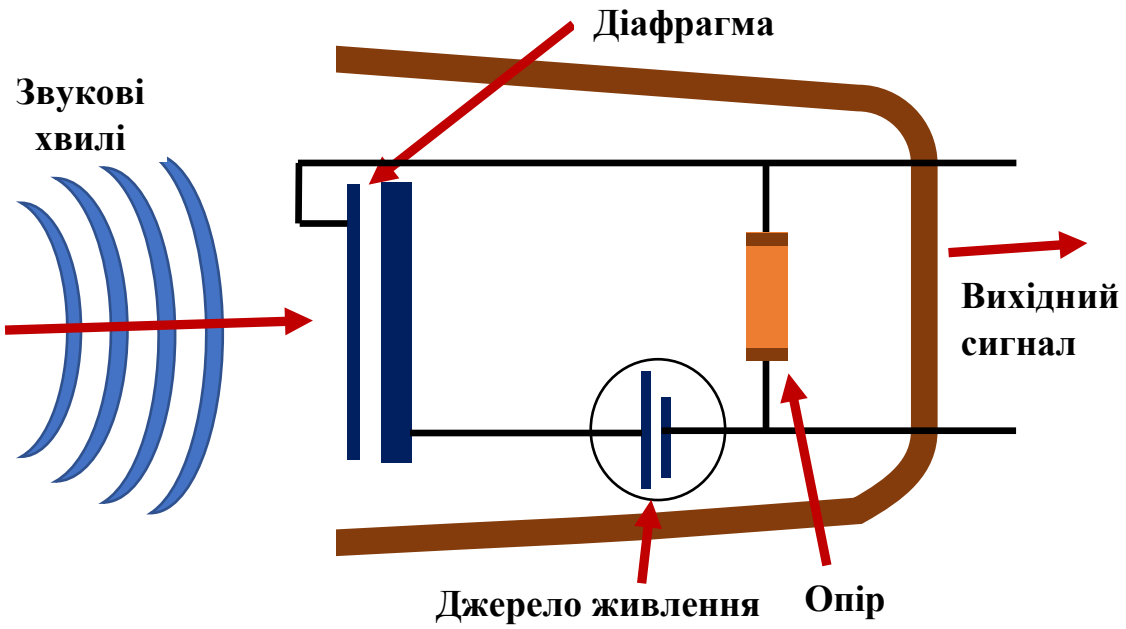


Рисунок 2.3 — Конструкція конденсаторного мікрофона

Електродинамічні мікрофони стрічкового типу були створені шляхом з’єднання мембрани та гофрованої стрічки з алюмінієвої фольги, товщиною приблизно 2 *мкм*, у магнітному полі (рис. 2.4). Генерація електричних сигналів

відбувається за рахунок коливань цієї стрічки. Основною перевагою таких мікрофонів є надзвичайно широкий спектр частотних характеристик, але їхня чутливість доволі невелика, вихідний опір долі *Ом*, що зі свого боку сильно ускладнює проектування підсилювачів для них. Необхідного рівня чутливості можна досягти, лише через збільшенням площі стрічки (відповідно й розмірів магніту), але тоді такі мікрофони будуть мати велику масу та розміри в порівнянні з іншими типами.

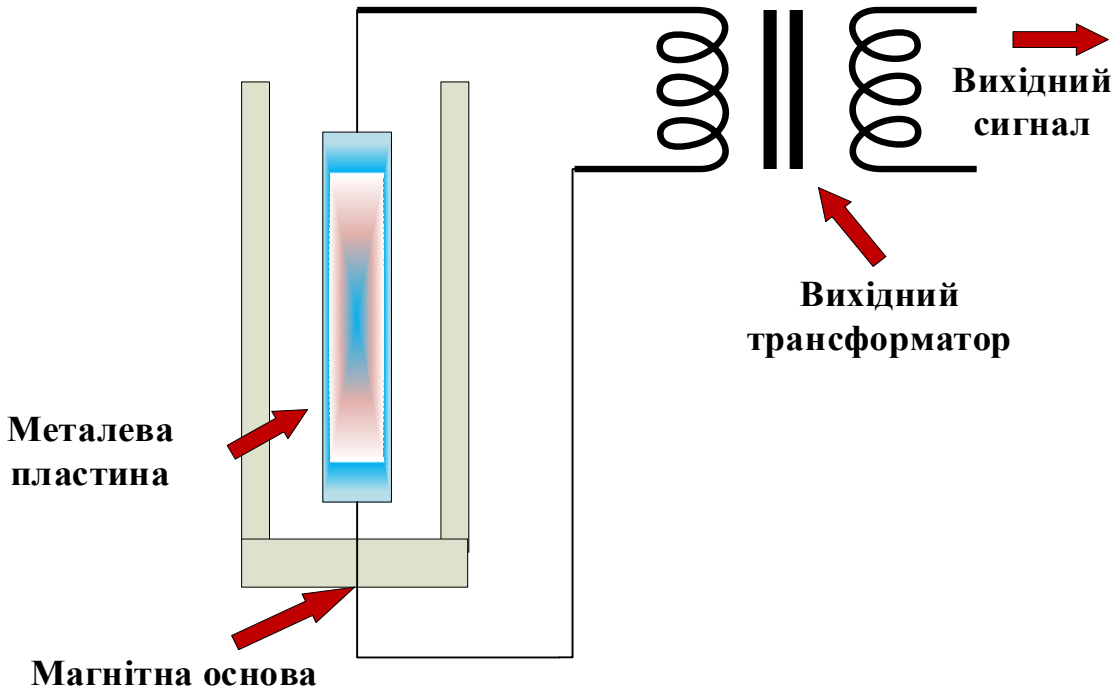


Рисунок 2.4 — Конструкція динамічного мікрофона стрічкового типу

Інший тип динамічних мікрофонів складається з тонкої мембранної пластини з фольги або полістиролу, до якої прикріплена вже не стрічка, а котушка, що знаходиться в магнітному полі (рис. 2.5). Під дією звукових хвиль мембрана та котушка починають коливатись, що в свою чергу породжує електричні сигнали. Такі мікрофони відрізнялися від стрічкових більш високим вихідним опором (від десятків *Ом* до сотень *кОм*), тому можуть мати менші розміри та є зворотним.

На відміну від електродинамічних мікрофонів, електромагнітні мають у своїй конструкції нерухому котушку, а магніт постійно закріплений на мембрані. У зв’язку з відсутністю характерних для динамічних мікрофонів жорстких вимог

щодо маси котушки, електромагнітні мікрофони більш універсальні та високоомні, завдяки багатовідвідним котушкам. Вони та п'єзоелектричні мікрофони використовувалися для створення ефективних слухових апаратів.

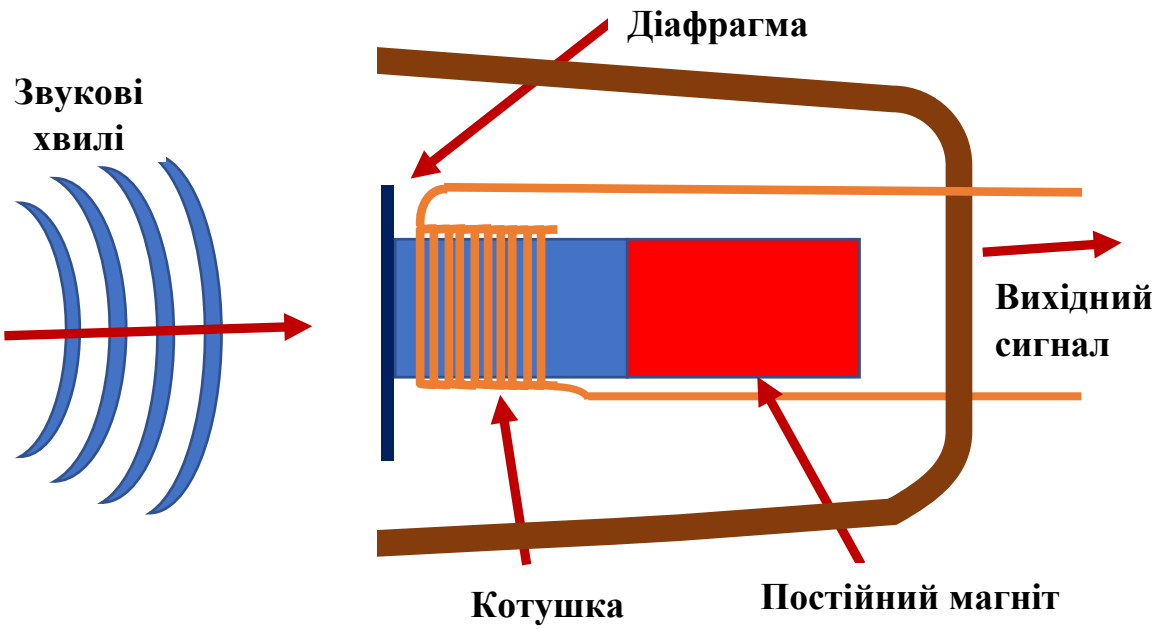


Рисунок 2.5 — Динамічний мікрофон котушкового типу

Електретні мікрофони за своєю конструкцією та принципом дії доволі схожі на конденсаторні мікрофони, тільки в ролі нерухомої обкладки конденсатору та джерела постійної напруги в них виступає пластина з електрета — речовини, що має здатність досить тривалий час бути поляризованою та створювати електромагнітне поле. Електретні, як і конденсаторні мікрофони, мають дуже високий вхідний опір (від одиниць *МОм*), тому разом із ними використовуються польові транзистори. Саме через їхнє використання почали з’являтися легкі, малі за розміром та не менш ефективні електретні мікрофони.

Не залежно від типу, усі мікрофони можна оцінити за наступними параметрами та характеристиками [16]:

1. *Чутливість;*
2. *Амплітудно-частотна характеристика;*
3. *Акустична характеристика мікрофона;*

4. *Характеристика направленості;*
5. *Рівень власних шумів мікрофона;*
6. *Динамічний діапазон;*
7. *Внутрішній опір.*

Чутливість мікрофона — співвідношення напруги U на виході мікрофона до звукового тиску $p_{зв}$ у вільному акустичному полі, тобто за відсутності сигналу. А *осьовою чутливістю* називається шлях розповсюдження акустичної хвилі в напрямку акустичної осі мікрофона:

$$M_0 = \frac{U}{p_{зв}} \quad (\frac{мВ}{Па})$$

Вісь симетрії мікрофонів збігається з акустичною віссю. Якщо ж конструкціями мікрофонів не передбачено осей симетрії, то зазвичай у технічних умовах вказують напрямок акустичної осі. Також чутливість залежить від параметрів фізичних процесів у акустично-електричних перетворювачах та площі мембрани мікрофону. Зі збільшенням площі мікрофонної мембрани, у квадратичній залежності зростає і рівень чутливості. Варто зауважити, високий рівень чутливості ще не показує високий рівень якості передачі звукових сигналів, а лише надає можливість мікрофону «чути» слабкі звуки.

Залежність осьової чутливості мікрофону від частоти акустичних коливань називають ***амплітудно-частотною характеристикою*** (рис. 2.6). Її нерівномірність вимірюються в *дБ*, як співвідношення чутливості мікрофонів на якійсь обраній частоті до чутливості на середніх частотах, наприклад, 1 кГц.

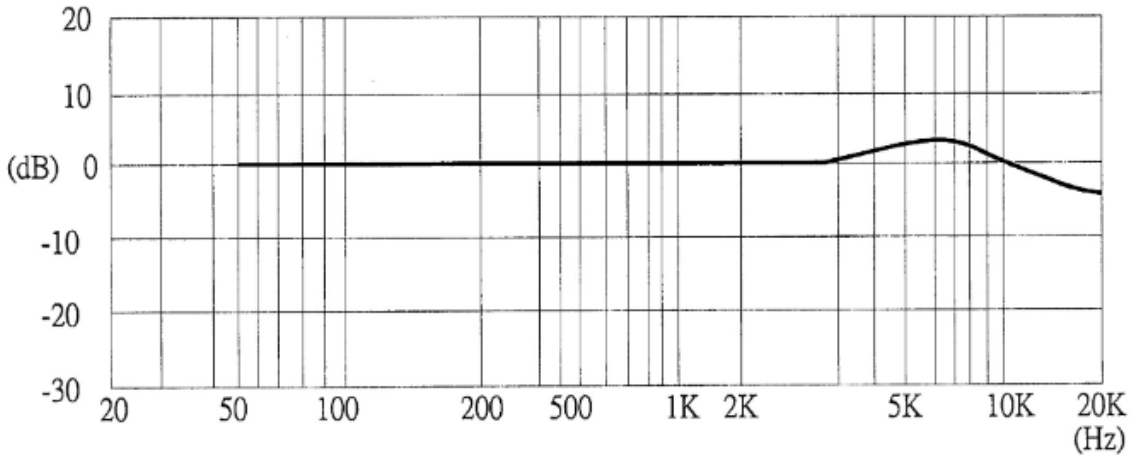


Рисунок 2.6 — АЧХ електретного мікрофона CMA-4544PF-W [21]

Акустична характеристика — співвідношення сили впливу акустичного поля, що дії на діафрагму мікрофонів *F*, до звукового тиску у вільному акустичному полі p_0:

$$A = \frac{F}{p_0}$$

З огляду на це, чутливість мікрофону *M* можна виразити через акустичну характеристику:

$$M = A \cdot \frac{U}{F}$$

Мікрофони, як електромеханічні перетворювачі, характеризуються множником *U/F* — співвідношенням вихідної напруги мікрофону та сили, яка діє на його діафрагму.

За акустичними характеристиками мікрофонів також визначаються і ***характеристики*** їхньої ***направленості*** — чутливість мікрофонів під час падінні на них звукових хвиль під кутом α відповідно до їхніх акустичних осей M_α поділена на осьову чутливість M_0 :

$$\varphi = \frac{M_\alpha}{M_0}$$

Завдяки характеристиці направленості мікрофону можна визначити місце розташування джерел акустичних сигналів. У випадку, коли φ = 1, тобто рівень чутливості не залежить від кута падіння акустичних хвиль, мікрофони називають *ненаправленими*. Такі види мікрофонів здатні приймати акустичні сигнали від джерел, що розташовані навколо нього.

Якщо ж значення чутливості мікрофону змінюється залежно від кута приймання акустичних сигналів, то для визначення джерел звуку, об’єкти мають бути розташовані в просторовому куті, у межах якого, значення чутливості мікрофону приблизно дорівнює осьовій чутливості.

За видами характеристик направленості можна виділити 3 основні типи мікрофонів: *приймачі тиску*, *градієнту тиску* та *комбіновані*.

Ненаправлені мікрофони або *приймачі тиску* — це такі види мікрофонів, які не мають направленої дії на частотах нижче 1 *кГц*. У них звуковий тиск біля

поверхні діафрагми визначається силою дії акустичних хвиль на діафрагму. Через свою специфічну конструкцію, одна сторона діафрагми захищена від дії акустичного поля, а інша — ні. Такі мікрофони можуть не змінювати звукове поле, якщо будуть мати розміри менші, за довжину акустичної хвилі, а діаграма направленості матиме вигляд кола (рис. 2.7 а). Якщо ж розміри будуть більшими, тоді за рахунок дифракції хвиль зміниться тиск та мікрофон стане направленим.

Приймачі градієнта тиску або *мікрофони двостороннього направлення* — це такі види мікрофонів, у яких поле акустичних коливань діє на обидві сторони діафрагми. Діаграма направленості таких мікрофонів зображена на рисунку 2.7 б та має вигляд «вісімки». На рухомі системи таких мікрофонів зазвичай діє сила, що визначається, як різниця звукових тисків з обох сторін діафрагми.

Мікрофонами *комбінованого типу* називаються односторонньо направлені. Форма їхньої діаграми направленості має вигляд кардіоїди (рис. 2.7 в), тому їх часто називають кардіоїдними.

Кардіоїдні мікрофони з більш вузькою направленістю називають суперкардіоїдними та гіперкардіоїдними (рис. 2.7 г). Основним недоліком цих різновидів мікрофонів є те, що на відміну від кардіоїдних, вони чутливі до сигналів з іншої сторони діафрагми.

Головна перевага під час використанні таких мікрофонів — розташування з одної сторони мікрофону, у межах широкого просторового кута, джерела звуку, якщо ж звукові хвилі розповсюджуються за межами просторового кута, мікрофон їх не сприймає.

Динамічним діапазоном називається різниця між граничним значенням звукового тиску та рівнем власних шумів мікрофона (максимальним і мінімальним сигналом), за якої вихідний сигнал знаходиться в допустимих межах.

Внутрішній опір Z_i для 90% мікрофонів є суто активним та частото-незалежним. В іншому випадку, якщо Z_i залежить від частоти, у технічній документації вказують середнє значення або модуль опору на частоті 1 кГц.

Внутрішній опір характеризує вимоги до вхідного опору пристрою (зазвичай — підсилювача), до якого під’єднано мікрофон.

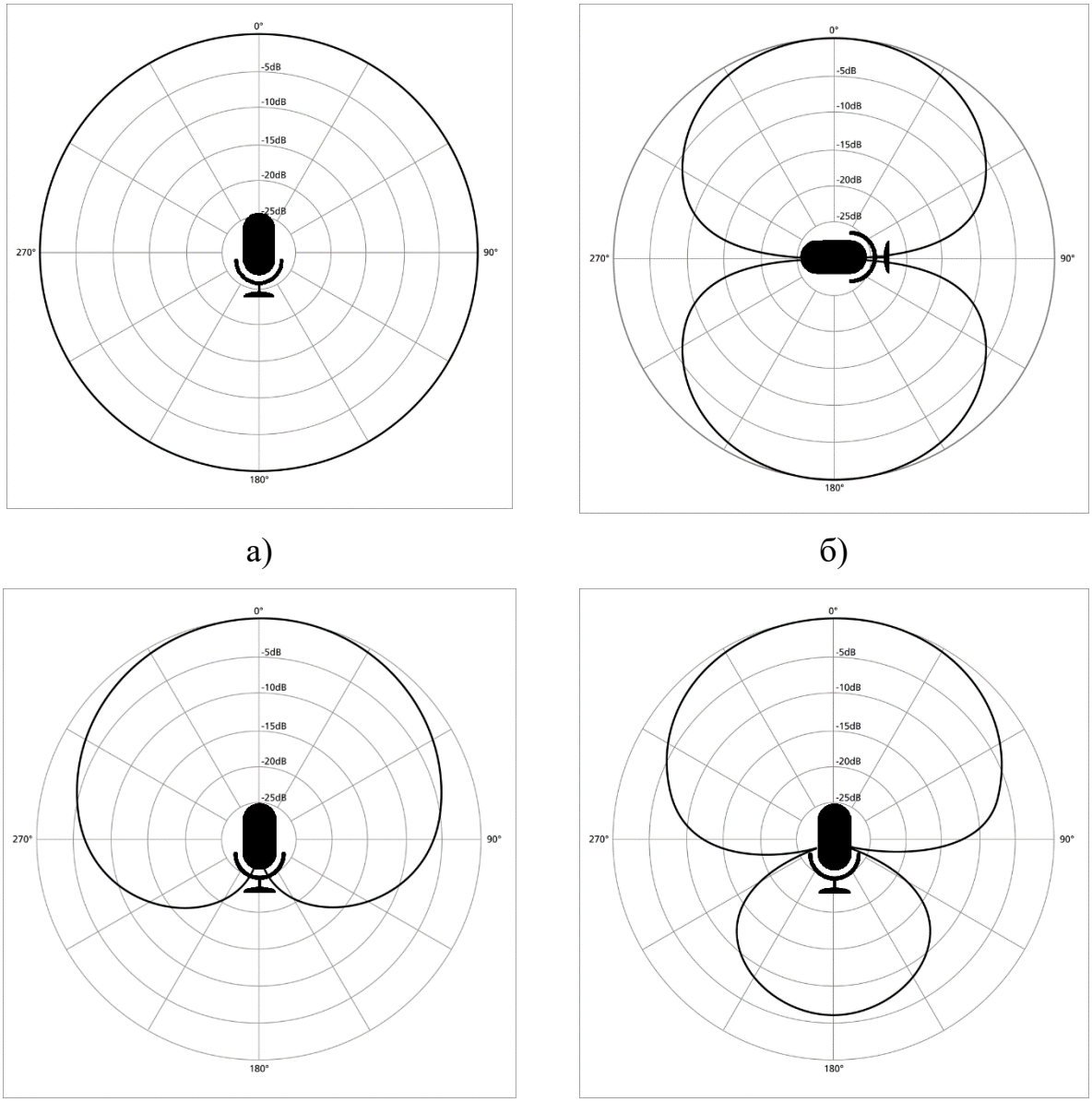


Рисунок 2.7 — Діаграма направленості: а) мікрофонів-приймачів тиску, б) мікрофонів-приймачів градієнтного тиску, в) комбінованих кардіоїдних мікрофонів, г) комбінованих гіперкардіоїдних мікрофонів.

Рівень власних шумів мікрофона або ***рівень еквівалентного звукового тиску*** $L_ш$ — співвідношення вихідної напруги мікрофону $U_ш$, що створюється ефективним звуковим тиском , до напруги U_1, що виникає в ньому тільки завдяки власним шумам за відсутності звукових коливань:

$$L_{ш} = 20 \cdot lg \frac{U_{ш}}{M \cdot p_0}$$

де $U_ш$ — корінь квадратний із різниці квадратів вихідної напруги на випробувальному стенді згідно з ГОСТ 16123-88 (IEC 60268-4), що виміряна спочатку з підключеним мікрофоном, потім, замість мікрофону, підключеним резистором – еквівалентом модулю опору випробуваного мікрофона, M – чутливість мікрофона на частоті 1 *кГц*, $p_0 = 2 \cdot 10^{-5}$ *Па*.

Міжнародні стандарти на мікрофони визначають такі параметри: *еквівалентний рівень шуму* та *співвідношення «сигнал/шум»*. У різних стандартах методи їхнього вимірювання відрізняються, тому в сучасних каталогах вказують два значення рівня шумів відповідно до стандарту. Наприклад, для конденсаторного мікрофона C-3000B фірми AKG еквівалентний рівень шумів за стандартом IEC-651 (DIN45412) дорівнює 14 дБ-А ("А" означає метод моделювання сприйняття людиною), а за другим стандартом CCIR468-2 (DIN45405) — 25 дБ. А співвідношення «сигнал/шум» складає 80 дБ (DIN / IEC-651) і 69 дБ (CCIR 468-2) [22].

Окрім наведених вище параметрів, для мікрофонів також важливими є ***робочий діапазон зміни температури*** та ***вологість навколишнього середовища***, всередині якого рівень чутливості не має змінюватися більше, ніж на ± 2 дБ. Діапазон температур для стабільної роботи всіх груп електродинамічних мікрофонів знаходиться в межах від -40°С до +50°С, а максимальна вологість не має перевищувати 95% за температури 20°С. Для конденсаторних мікрофонів за такої ж температури вологість має бути не більше 85 %, а діапазон робочих температур — від -10°С до + 35°С.

Зараз найбільш ефективними та популярними у використанні є електродинамічні (котушкового та стрічкового типу) і електростатичні (конденсаторні та електретні) мікрофони.

Основними ***перевагами*** динамічних мікрофонів є:

- *Менша чутливість до відлуння сигналів*, ніж у конденсаторних мікрофонів;

- *Стійкі до високого звукового тиску* ;
- *Надійні* — високий рівень стійкості до механічних пошкоджень, екстремальних робочих температур та вологості;

З ***недоліків*** можна виділити таке:

- *Мала чутливість, слабша реакція* на перехідні процеси та *обмежена частота реєстрації* акустичних сигналів;

Через це динамічні мікрофони зазвичай фіксують менше слабких акустичних сигналів, ніж конденсаторні мікрофони.

- *Високий рівень нерівномірності АЧХ;*
- *Великі розміри* та *маса* необхідні для досягнення високого рівня чутливості;
- *Висока ціна* на якісні моделі мікрофону.

Конденсаторні мікрофони мають наступні ***переваги*** та ***недоліки***:

- *Висока чутливість* та *швидкість реакції* на перехідні процеси.

Через це конденсаторні мікрофони мають здатність записувати тихі сигнали, деталі та зміни тону звуку.

- *Широкий частотний діапазон;*
- *Мала нерівномірність АЧХ;*
- *Низький рівень шумів;*
- Необхідно *«фантомне» живлення 48 В*;
- *Висока ціна* на якісні моделі мікрофону.
- *Крихкі* — мають високу схильність до механічних пошкоджень;
- *Низькій рівень стійкості до несприятливих погодних умов* (екстремальних робочих температур та опадів);

Через що більшість конденсаторних мікрофонів малоефективні під час використанні на свіжому повітрі без спеціального захисного обладнання.

Найрозповсюдженіший типом мікрофону є електретний, що має наступні ***переваги*** [23]:

- *Широкий частотний діапазон;*
- *Мала нерівномірність АЧХ;*

- *Висока чутливість* та *якість* звукозапису;
- *Малі габаритні розміри;*
- *Низька ціна ;*
- *Надійні* — високий рівень стійкості до механічних пошкоджень, екстремальних робочих температур та вологості;

Недоліком є те, що електретні мікрофони, як і конденсаторні, потребують зовнішнього *«фантомного» живлення*. Для більшості видів мікрофонів це значення невелике, приблизно 3 *В,* що цілком можна забезпечити використанням акумулятора або через кабель mini-Jack 3.5 мм.

Розглянувши наявні на даний момент на ринку моделі мікрофонів, у таблицю 2.1 були внесені апроксимовані середні значення їхніх основних характеристик та класифіковані за типом мікрофонів:

Таблиця 2.1 — Порівняльні характеристики основних типів мікрофонів

Тип мікрофону	**Електродинамічний**		**Конденсаторний**	**Електретний**
	котушковий	**стрічковий**		
Діапазон частот, *Гц*	50 – 15 000 (1 клас) 40 – 20 000 (вищий клас)	50 – 15 000 (1 клас) 30 – 18 000 (вищий клас)	40 – 18 000 (1 клас) 20 – 20 000 (вищий клас)	40 – 18 000 (1 клас) 20 – 20 000 (вищий клас)
Чутливість, *дБ*	-60 … -45	-56 … -40	-60 … -30	-60 … -30
Нерівномірність АЧХ, *дБ*	12	10	5	5
Рівень власних шумів, *дБ*	30 … 18	30 … 16	30 … 4	30 … 5
Внутрішній опір, *Ом*	150 … 1500	200 … 1000	200 … 2200	200 … 2200
Звуковий тиск, *дБ*	до 150	до 140	до 140	до 140

За даними з таблиці 2.1 проведемо порівняльний аналіз усіх наведених типів мікрофонів та виберемо найбільш підходящий для подальшого дослідження.

Чутливість — найважливіший параметр для оцінювання мікрофону, з таблиці 2.1 випливає, що найбільший рівень *чутливості* мають конденсаторні та електретні типи мікрофонів — -30 *дБ*. Оскільки, величини чутливості мають

від'ємні значення, то потрібно керуватися такими правилами, що чим ближче це значення до 0, тим більш чутливим буде мікрофон.

Нерівномірність амплітудно-частотної характеристики показує в скільки разів або на скільки *дБ* потужність прийнятого сигналу (за амплітудою) зменшується на заданому частотному діапазоні. В ідеальному випадку, графік АЧХ мусить бути рівномірним на всьому діапазоні робочих частот мікрофону, тобто за своєю формою наближатися до горизонтальної лінії. Зараз у вимірюванні акустичних сигналів активно застосовуються конденсаторні та електретні мікрофони, через їхній малий коефіцієнт нерівномірності амплітудно-частотної характеристики.

Діапазон частот. Корисні гармоніки спектру акустичних сигналів малих безпілотних літальних апаратів лежить у проміжку від 1 кГц до 12 кГц. Цей проміжок входить у загальний *динамічний діапазон робочих частот* усіх наведених типів мікрофонів, тому за цим параметром підходить будь-який із них для детектування сигналів від дронів.

Звуковий тиск, що може сприймати мікрофон — максимальне значення рівня гучності звуку, за якого не виникає помітний вплив завад на прийнятий акустичний сигнал. Чим більше значення звукового тиску, тим вище якість роботи мікрофона з гучними звуками. Електродинамічні, конденсаторні та електретні мікрофони найчастіше здатні працювати в разі максимального звукового тиску 140 дБ. Згідно з класифікацією загальних рівнів звукового тиску, цей рівень еквівалентний шуму під час зльоті реактивного літака на відстані пари метрів або від пострілів із вогнепальної зброї [24]. А граничним значенням роботи динамічних мікрофонів є рівень 150 – 160 дБ, це звуки, наприклад, вибухів ракет або артилерійських снарядів. Шум за такого рівня нестерпний для людини та завдає сильні фізичні ушкодження. Динамічні мікрофони є безумовними лідерами на ринку за максимальним рівнем звукового тиску, але й показники інших типів мікрофонів знаходяться в межах, прийнятних для дослідження, оскільки рівень випромінювання шумів нинішніх МЛА на потрібних для детектування відстанях не вище 90 дБ.

<u>Внутрішній опір</u> або «*імпеданс*» — здатність мікрофона чинити опір змінному електричному струму. Він показує чи сумісним є мікрофон із підсилювачем, щоби під час роботи не відбувалися втрати потужності сигналу.

Динамічні мікрофони зазвичай мають мале значення імпедансу, а конденсаторні та електретні — високе. Від значення імпедансу залежить вид конструкції підсилювача.

<u>Рівень власних шумів</u> показує можливості мікрофона працювати зі слабкими акустичними сигналами, чим менше значення рівня, тим тихіший акустичний сигнал може чітко передавати мікрофон.

Діапазон рівнів сигналу на виході мікрофону зазвичай складає від *100 мВ* до *1 В*, тому відносний рівень власних шумів усієї системи визначається шумами підсилювача або передпідсилювача-коректора [25]. У пристроях із динамічними мікрофонами, імпеданс яких лежить у межах від 200 Ом до 2 кОм, вхідні каскади підсилювачів будуються на біполярних транзисторах, а вхідні каскади підсилювачів для конденсаторних та електретних мікрофонів із вихідною ємністю $C_{дж} = 5...40$ *пФ* — на польових транзисторах [26].

Власні шуми *<u>динамічних мікрофонів</u>* легко проаналізувати за допомогою еквівалентної шумової моделі (рис. 2.8). Більшою мірою шуми утворюються підсилювачем на біполярному транзисторі, тому ЕРС шуму дротів $\varepsilon_{ш.др}$, індуктивністю обмотки котушки L та міжвитковою ємністю C можна знехтувати.

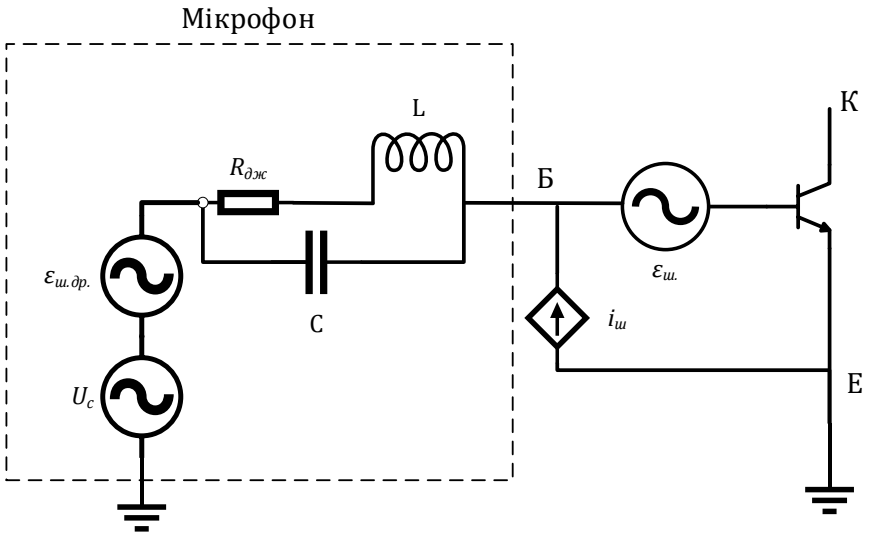


Рисунок 2.8 — Еквівалентна шумова схема системи на біполярному транзисторі

Отже, повна напруга шумів системи $\varepsilon_{\text{ш.заг}}$ дорівнює:

$$\varepsilon_{\text{ш.заг}} = \sqrt{\varepsilon_{\text{ш}}^2 + i_{\text{ш}}^2 R_{\text{дж}}^2} \qquad \left(\text{В}/\sqrt{\text{Гц}}\right)$$

Перший доданок у скобках — вхідна ЕРС шуму, другий — напруга шуму, що утворюється проходженням вхідного струму шуму дротами динамічного мікрофону, що мають значення опору $R_{\text{дж}}$. Проаналізуємо основні параметри доданків більш детально.

Шум напруги $\varepsilon_ш$. Еквівалентний генератор напруги шуму розглядають, як ввімкнений послідовно з базою транзистора. Генератор представляє собою суму теплового шуму, що виникає через хаотичний рух зарядів, який призводить до падіння напруги на опорі бази транзистора $r_б$, та дробового шуму, що генерує шум напруги на диференційному опорі емітера $r_е$.

Наведені теплові шуми аналогічні тепловим шумам резисторів, тому їхню спектральну щільність можна визначити за формулою Найквіста: $e_{\text{т.ш}} = \sqrt{4kTr_{\text{б}}}$, де $k = 1.38 \cdot 10^{-23}$ Дж/К – стала Больцмана; T – абсолютна температура опору $r_б$, K. А спектральна щільність дробового шуму описується виразом $i_{\text{д.ш}} = \sqrt{2qI_{\text{к}}} \cdot r_e$, де $I_к$ – постійний струм колектору.

Відповідно шум напруги дорівнює:

$$e_{\text{ш}}^2 = 4kTr_{\text{б}} + 2qI_{\text{к}}r_{\text{е}}^2 = 4kTr_{\text{б}} + 2(kT)^2/\, qI_{\text{к}} \quad (\text{В}^2/\text{Гц})\ .$$

Також через нерівномірність процесу рекомбінації носіїв на поверхні напівпровідників, виникає явище флікер-шуму. Проте він має суттєвий ефект тільки на низьких частотах та за великих значень струму бази, тобто під час великого значення струму колектору, відповідно $e_ш$ постійне у великому діапазоні значень струму колектору; воно збільшується за малих значеннь струму (дробовий шум струму крізь зростаючий опір $r_е$) та за достатньо великих струмах (флікер-шум від проходження $I_б$ крізь $r_б$)

Шум струму $i_ш$. Основним джерелом шуму струму є флуктуації дробового шуму в струмі бази, що додаються з флуктуаціями за рахунок флікер-шуму у $r_б$. Для біполярного транзистора вплив флікер-шуму визначається введенням

додаткового множника $\sqrt{1+\frac{f_ф}{f}}$ у вираз дробового шуму струму бази $i_ш = \sqrt{2qI_б(1+\frac{f_ф}{f})}$, де $f_ф$ – частота зрізу флікер-шумів, на якій спектральна густина зростає на 3 дБ по відношенню до спектральної густини на високих частотах. Для спеціальних малошумлячих транзисторів $f_ф$ = 200…500 Гц, а для транзисторів загального використання $f_ф$ = 10…40 кГц.

На рисунку 2.9 наведені апроксимовані графіки залежності вхідної напруги шуму $e_ш$ та вхідного струму шуму $i_ш$ від струму колектору $I_к$ та частоти f для біполярних транзисторів. З графіків випливає, що $e_ш$ спадає, а $i_ш$ зростає зі збільшенням $I_к$, це дає можливість оптимізувати робочий струм транзистора для отримання мінімального шуму у разі заданого джерела сигналу. Щоб зменшити повний шум підсилювача $e_{під}$, потрібно знати $R_{дж}$, далі із залежності $e_ш$ та $i_ш$ від $I_к$ на частотах сигналу вибрати $I_к$, що мінімізує $e_{під}$ (рис. 2.10).

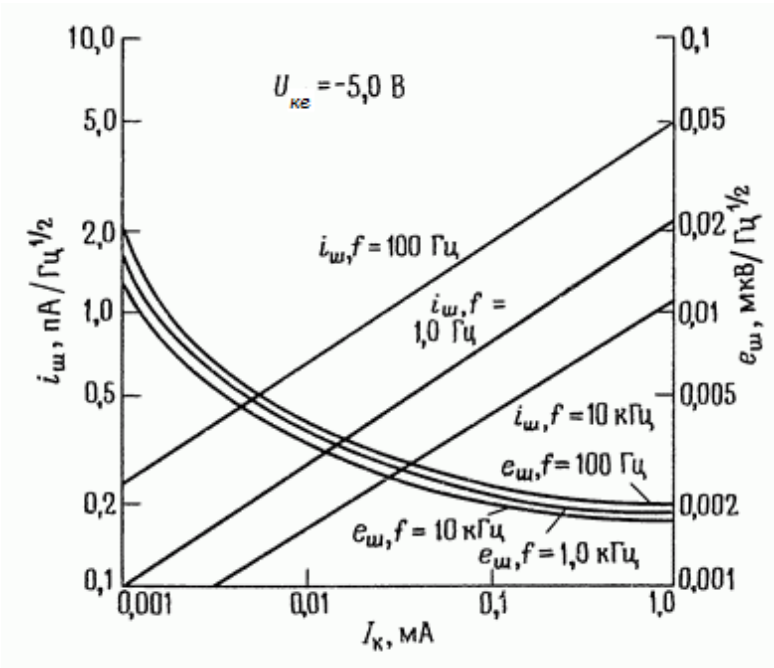


Рисунок 2.9 — Залежність еквівалентної середньоквадратичної $e_ш$ та $i_ш$ від $I_к$ для біполярного транзистора 2N5087 [27]

У підсумку зауважимо, що на низьких частотах (*до 5 кГц*) та за малого імпедансу джерела вхідного сигналу (до *1 кОм*) завжди будуть переважати шуми напруги біполярного транзистора, тому середнє виміряне значення рівня власних шумів для динамічних мікрофонів дорівнює приблизно 16 *дБ*.

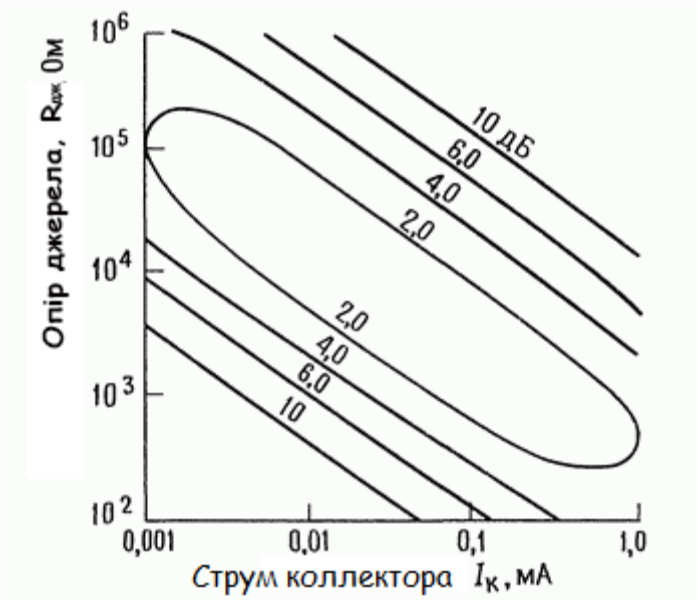
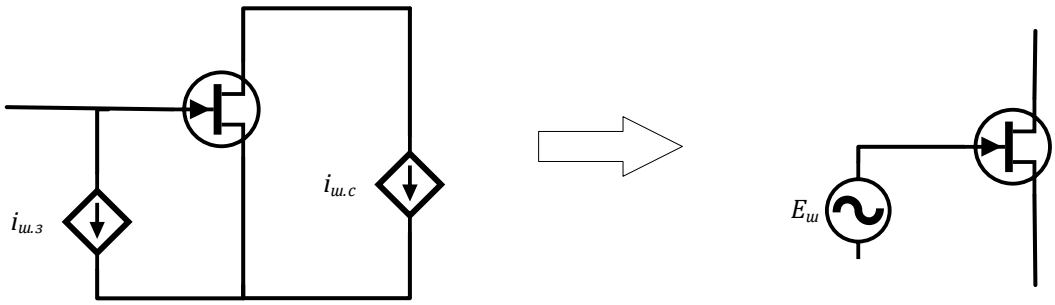


Рисунок 2.10 — Лінії рівня коефіцієнта шуму для біполярного транзистора 2N5087 за умови f = 1 кГц та ширині смуги 150 Гц [27]

На вході *конденсаторних* та *електретних* мікрофонів стоїть транзисторний перетворювач заряд-напруга, тому через високий вхідний опір рівень власних шумів загалом буде визначатися польовим транзистором, що розташований безпосередньо у капсулі мікрофону.

Зазвичай є два основних джерела струму для польового транзистора з p-n-переходом (рис. 2.11 а). Генератор шумового струму затвора $i_{ш.з}$. характеризує дробовий шум зворотно зміщенного p-n-переходу, $i_{ш.з} = \sqrt{2qI_з}$, де $I_з$ – струм заслону. Тепловий рух електронів у каналі викликає шуми стоку $i_{ш.с.}$, спектральна щільність яких $e_{ш.с} = \sqrt{4kTq_к}$, де $q_к$ — ефективна провідність каналу , що пов'язана з крутизною транзистора S залежністю $q_к = 0.7S$.



а) **б)**

Рисунок 2.11 — Еквівалентна шумова схема польового транзистора

Генерація процесу рекомбінації на поверхні p-n-переходу є джерелом виникнення флікер-шумів у польовому транзисторі. Як і теплові, ці шуми модулюють ширину каналу та викликають додаткові шуми в колі стоку. З урахуванням флікер-шуму та підставивши значення $q_к$, вираз для спектральної щільності струму стоку має вигляд $e_{\text{i.c}} = \sqrt{2.8kTS(1 + \frac{f_\text{ф}}{f})}$.

Середньоквадратичні значення $e_ш$ та $i_ш$ у діапазоні частот від $f_н$ до $f_в$ для польового транзистора:

$$e_\text{ш} = \sqrt{2.8kTS[f_\text{в} - f_\text{н} + f_\text{ф} \ln\left(\frac{f_\text{в}}{f_\text{н}}\right)]}, \qquad i_\text{ш} = \sqrt{2qI_\text{з}\,(f_\text{в} - f_\text{н})}$$

Струми затвору польових транзисторів одиниці пА, тому падіння шумової напруги на внутрішньому опорі джерела, що спричинене струмом генератора $i_ш$, достатньо мале й на практиці прийнято вважати рівним нулю. Отже, польовий транзистор, на відміну від біполярного, описується шумовою еквівалентною схемою з одним джерелом шумової ЕДС $e_ш$, яка ввімкнена послідовно в ланку затвору (рис. 2.11 б). Шум напруги $e_ш$ може бути зменшений шляхом збільшення I_c до такого значення, за якого крутизна буде найбільшою.

Власні шуми каналу польових транзисторів із ізольованим затвором (МОП) мають один порядок з шумами польових транзисторів з p-n-переходом розглянутими вище. Однак через те, що оксид має некристалічну структуру та дефекти, МОП транзистори неможливо використовувати на вході досить чутливих підсилювачів низьких частот, через значний рівень флікер-шумів.

Підсумовуюче вище сказане, коефіцієнт шуму в польових транзисторах значно менше, ніж у біполярних, оскільки струм транзистора утворюється носіями заряду основного типу, концентрація яких мало залежить від температури та процеси рекомбінації не відіграють у них значної ролі. Головними складовими шуму польових транзисторів є теплові шуми, що виникають у провідному каналі, і дробові шуми струму затвора.

Тому на практиці найнижчі показники рівня власних шумів зустрічаються, лише, у сучасних конденсаторних мікрофонах із великою діафрагмою та

електретних мікрофонах через схожість вхідних схем їхньої внутрішньої конструкції. Середньоквадратичне виміряне значення коефіцієнта власних шумів для цих мікрофонів дорівнює 5 *дБ*.

На основі проаналізованих вище параметрів, залишаємо для подальшого розгляду конденсаторні та електретні мікрофони. Вони мають схожий принцип роботи, однакого високий рівень чутливості та якості передачі звуку, через малий коефіцієнт власних шумів. Однак конденсаторному мікрофону потрібна вища напруга живлення 48 В. Потрібно зауважити, що під час детектування сигналів у польових умовах за високого рівня вологості, напруга в конденсаторному мікрофоні вище 40 В призводить до пробою. Тому для подальшого дослідження краще використовувати електретні мікрофони, по-перше, через нижчу напругу живлення; по-друге, високу надійність — вони більш стійкі до високих температур та рівня вологості середовища; по-третє, такі моделі мікрофонів дешеві.

У якості приймального пристрою в прототипі акустичного детектора малих безпілотних літальних апаратів обрано електретний мікрофон ***CMA-4544PF-W***, що має наступні параметри [21]:

- Чутливість: -44 ± 2 дБ (на частоті 1 кГц та тиску 1 Па);
- Амплітудно-частотна характеристика зображена на рисунку 2.6;
- Характеристика направленості — всенаправлений;
- Співвідношення сигнал / шум: 60 дБ;
- Діапазон частот: 20 … 20000 Гц;
- Внутрішній опір: 2,2 кОм;
- Напруга живлення: 3 … 10 В;
- Робоча температура: -20 … +70 °C.

2.2. Проектування підсилювача до мікрофона. Розподіл фільтрації між акустичним трактом та підсилювачем.

Напруга сигналу на виході мікрофонів зазвичай має досить низький рівень, від 0.2 до 2 мВ, тому при використанні конденсаторного або електретного

мікрофону буде достатньо коефіцієнту підсилення 50 – 100. Схеми мікрофонних підсилювачів мало відрізняються від звичайних підсилювачів низьких частот (ПНЧ). Для них залишається справедливою вимога малих нелінійних спотворень (приблизно 1 – 2%). Для зменшення нелінійних спотворень у мікрофонних підсилювачах зазвичай застосовують негативні зворотні зв'язки.

Досить важливою вимогою до мікрофонного підсилювача є повна відсутність фону змінного струму. Наявність фону призводить, по-перше, до шуму сигналу, по-друге, до появи з обох боків від подавленої несучої двох частот, що відрізняються від неї на 50 або 100 Гц. Для знищення фону використовують фільтрацію напруги живлення, продумане розміщення деталей та проводів, підбір точки заземлення деталей першого каскаду, екранування проводів та деталей.

Мікрофонні підсилювачі нерідко працюють в умовах сильних високочастотних полів, які можуть спричинити наведення високочастотної напруги на мікрофоні, мікрофонних шнурах та деталях підсилювача. На вході вона може досягати одиниць або десятків вольт, що спричиняє появу сіткового струму, різке порушення режиму роботи ПНЧ та дуже великі спотворення. Може виникнути також самозбудження підсилювача на частотах від кількох одиниць Гц до десятків кГц, що спричиняють побічні випромінювання. Якщо мікрофон не екранований, його потрібно екранувати металевою сіткою чи ґратами. Часто допомагає окреме з'єднання мікрофона із заземленням або корпусом.

Мікрофонний підсилювач повинен мати наступні ***параметри***:

- Діапазон частот 50 – 15000 Гц;
- Нерівномірність амлітудно-частотної характеристики — не більше ±0,5 дБ на частоті 1 кГц;
- Вхідний опір мікрофонного підсилювача, при якому забезпечується максимальне співвідношення сигнал/шум на всьому діапазоні частот, вибирається втричі більше, ніж повний внутрішній опір мікрофону.
- Зміна модуля вхідного опору в залежності від частоти - не більше 5%;
- Коефіцієнт гармонік — не більше 0,5%;

- Напруга власного рівня шумів та фону на вході — не більше 5 мкВ.
- Великий динамічний діапазон

Однією з особливостей роботи реєстратора малих безпілотних літальних апаратів є те, що амплітуди акустичних сигналів дронів можуть змінюватись в дуже широкому діапазоні в залежності від відстані або оточуючого середовища. Це може викликати процес перевантаження приймача, тому що налаштування підсилювача, які необхідні для реєстрації мікрофоном ехо-сигналів цілей на малих відстанях, не підходять для прийому сигналів, відбитих цілями, що знаходяться на великій дальності та навпаки. Відомо, що середнє значення вихідної напруги мікрофона $U_{вих}$ зменшується зі зростанням дальності, а величина флуктуації залишається постійною. Для послаблення впливу завад, необхідно стиснути флуктуації до рівня власних шумів приймача.

Відповідно, щоб приймач не досягав рівня насичення при зміні вхідних сигналів у широких межах, тобто мав великий динамічний діапазон, зараз використовують *логарифмічні підсилювачі* або *системи автоматичного регулювання підсилення* (АРП) , які залежно від величини вхідного сигналу, поділяється на тимчасове, швидкодіюче та інерційне автоматичне регулювання підсилення [28].

Логарифмічний підсилювач — підсилювач, де залежність коефіцієнта підсилення від амплітуди вхідних сигналів описується логарифмічною функцією. Для сигналів низької амплітуди відбувається майже лінійне підсилення, а зі збільшенням амплітуди сигналів передаточна функція дедалі більше стає логарифмічною, відповідно слабкі сигнали зазнають більшого підсилення, а сильні — меншого.

Можливі різноманітні схемні реалізації для отримання логарифмічної характеристики у підсилювачі. Найбільш розповсюдження отримала схема з послідовно з’єднаних каскадів підсилення, кількість яких визначає динамічний діапазон підсилювача (рис. 2.12).

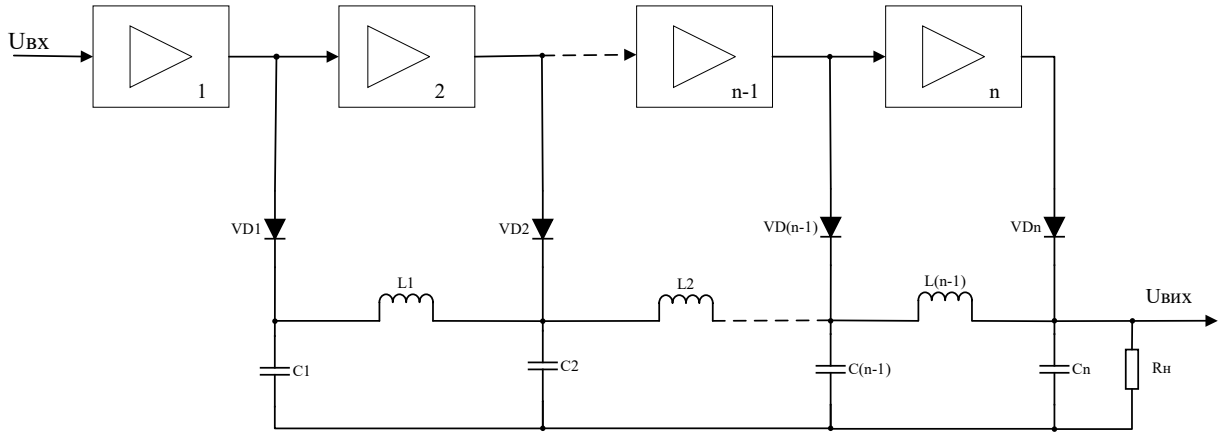


Рисунок 2.12 — Типова структурна схема логарифмічного підсилювача

Сумарний сигнал, який виділяється на загальному навантаженні $R_н$, з якого він далі надходить на диференціюючу ланку з малою сталою часу, як у лінійному підсилювачі частот.

Лінія затримки дозволяє всім імпульсам з виходу діодів VD_1-VD_5 приходити до навантаження одночасно (враховується затримка у кожному каскаді підсилювача). Амплітудна характеристика каскадів лінійна для малих амплітуд і має обмеження при якомусь значенні $E_{гр}$. Відповідно, імпульси великої амплітуди обмежуються і на суматор надходять з однаковою амплітудою, що дорівнює $E_{гр}$.

Вхідні імпульси різної амплітуди (зміни у великому діапазоні) обмежуються в різних каскадах підсилювача (найслабший – в останньому, найсильніший – у першому), за великої амплітуди відбувається менший приріст амплітуди вихідних імпульсів, ніж за малої. У результаті амплітудна характеристика набуває логарифмічної форми за рахунок складання окремих лінійних ділянок з нахилом, що поступово зменшується. Підбором коефіцієнтів підсилення каскадів та порогових значень для детекторів встановлюють необхідний вид передавальної характеристики підсилювача.

Після диференціюючої ланки, що має малу сталу часу, з вихідного сигналу підсилювача виключається постійна складова (видаляється середнє значення), і амплітуда шумів навколишнього середовища на будь-якій відстані буде на одному рівні з власним шумом приймача.

Також часто у логарифмічному підсилювачі у ланцюзі зворотного зв'язку замість діода використовують біполярний транзистор. Такий підхід забезпечує кращі характеристики та більший динамічний діапазон.

Основними ***перевагами*** використання логарифмічних підсилювачів є:

- *широкий робочий динамічний діапазон* — до 7 – 9 декад;
- *безінерційність*;

Такі підсилювачі здатні реагувати на регулярні та випадкові завади, можуть миттєво відновлювати чутливість після впливу сильних перешкод;

- *низький рівень шуму*;

Логарифмічні підсилювачі мають низькі шумові характеристики, що робить їх ідеальними для використання в додатках для підсилення сигналів низького рівня.

- нелінійний відгук;

Логарифмічна характеристика підсилювачів робить їх ідеальними для застосувань, де вихідна потужність повинна бути пропорційна логарифму вхідного сигналу, наприклад, в обробці аудіо або радіосигналів.

Головними ж ***недоліками*** логарифмічних підсилювачів є:

- *інтермодуляційна вибірковість*;

Через нелінійність характеристики підсилювача два слабких сигнали будуть давати додаткові «продукти» перетворення, тобто у спектрі почнуть з’являтись складові, яких у початковому сигналі не має.

- *обмежений частотний діапазон*;

Частотна характеристика логарифмічних підсилювачів може не підходити для деяких високочастотних застосувань, оскільки вони не можуть забезпечити достатню смугу пропускання.

- *обмежена точність*;

Хоча логарифмічні підсилювачі мають дуже високу чутливість, вони можуть не забезпечувати такий же рівень точності, як інші типи підсилювачів.

- *нелінійна реакція*;

Хоча логарифмічна характеристика підсилювача може бути перевагою, вона також може бути недоліком у деяких випадках, коли потрібна лінійна характеристика.

- *складний схемотехнічний дизайн*;

Конструкція логарифмічних підсилювачів складніша, ніж у інших підсилювачів, що може призвести до вищих виробничих витрат.

Підсилювачі з системою АРП використовують для більш ефективного придушення завад оточуючого середовища та отримання ширшого динамічного діапазону. Діапазон регулювання обмежений знизу рівнем власних шумів, а зверху — максимальною неспотвореною напругою. Точність регулювання залежить від коефіцієнта підсилення системи та обмежується її стійкістю. Глибина регулювання хорошого підсилювача з системою АРП повинна складати 40 – 100 дБ.

Працює система АРП наступним чином: при зміні амплітуди вхідного сигналу $U_{ВХ}$, змінюється коефіцієнт підсилення (напруга регулювання), що дає змогу підтримувати стабільний рівень вихідної напруги $U_{ВИХ}$. Тобто для вхідних сигналів з великою амплітудою коефіцієнт підсилення будем меншим, ніж для слабких сигналів. При цьому, якщо приймаються сигнали різних амплітуд, підсилення підлаштовується під найменший з них.

Основні способи регулювання підсилення в залежності від схеми [29]:

- режимне регулювання (зміна крутості прохідної характеристики транзистора $|Y_{21}|$ шляхом зміни режиму його роботи по постійному струму);
- зміна опору навантаження підсилювача;
- зміна глибини зворотного зв'язку;
- зміна величини міжкаскадного зв'язку за допомогою керованих атенюаторів;
- зміна ступеня зв'язку контуру з транзистором.

Основними ***перевагами*** використання мікрофонних підсилювачів з АРП є:

- *покращене співвідношення сигнал/шум*;

АРП допомагає збільшувати коефіцієнт підсилення для слабких сигналів і зменшувати для сильних, що зменшує шум і спотворення у вихідному сигналі.

- *покращена узгодженість*;

Регулюючи коефіцієнт підсилення аудіосигналу, АРП може допомогти в забезпеченні постійного рівня вихідного сигналу, незалежно від рівня вхідного сигналу.

- *лінійність*;

АРП може допомогти розширити динамічний діапазон аудіосигналів, зменшуючи спотворення, які можуть виникнути, коли рівень вхідного сигналу перевищує максимальний вихідний рівень підсилювача.

- *зменшення зворотного зв'язку*;

АРП може допомогти зменшити зворотний зв'язок, який виникає, коли підсилений аудіосигнал з гучномовця вловлюється мікрофоном, викликаючи небажану петлю звуку.

З ***недоліків*** можна виділити наступні:

- *обмежений динамічний діапазон*;

Мікрофонні підсилювачі з АРП мають обмежений динамічний діапазон. Якщо рівень вхідного сигналу перевищує максимальний вхідний рівень, схема АРП може бути не в змозі компенсувати підвищений рівень сигналу, що призведе до спотворення і відсікання вихідного сигналу.

- *підвищена складність*;

Мікрофонні підсилювачі з АРП складніші, ніж інші типи підсилювачів, що може призвести до збільшення вартості виробництва.

- *підвищений рівень шуму*;

Схеми АРП можуть вносити додатковий шум в аудіосигнал через природу контуру управління.

- *обмежений контроль*;

Використовуються схеми АРП із *прямим регулюванням,* зі *зворотним регулюванням* та *комбіновані схеми*.

Система АРП із прямим регулюванням на практиці не дає можливість досягти ідеальної характеристики регулювання (отримати постійну напругу на виході) та найвищу швидкість спрацювання. Вони не знайшли широкого практичного використання через складність реалізації конструкції каналу підсилювача. Також недоліком такої системи АРП є необхідність ввімкнення перед амплітудним детектором додаткового підсилювача з великим коефіцієнтом підсилення. На «пряму» АРП суттєво впливають температурні зміни, що збільшують коефіцієнт підсилення, тим самим нарощують вихідну напругу дестабілізуючи регулювання.

Частіше на практиці застосовуються АРП зі зворотним регулюванням (рис. 2.13). В залежності від рівня вхідного сигналу, напруга випрямляється амплітудним детектором, далі проходить крізь підсилювач постійного струму і ФНЧ та створює напругу регулювання. Ця напруга підводиться до регульованих каскадів підсилювача і зменшує коефіцієнт підсилення відповідно до регулювальної характеристики підсилювача. У результаті вихідна напруга підсилювача змінюється, відповідно до рівня вхідної напруги.

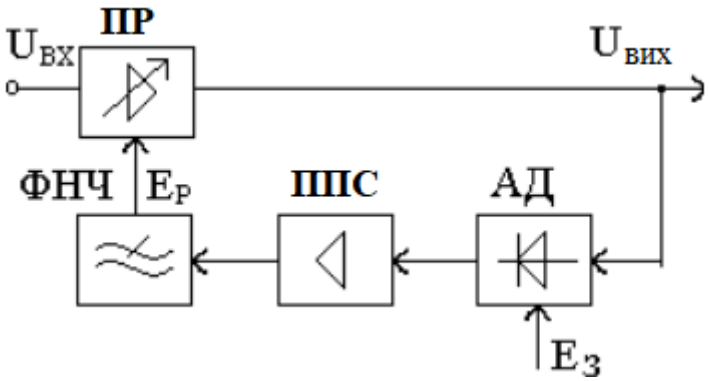


Рисунок 2.13 — Структурна схема системи АРП зі зворотним регулюванням, де ПР – підсилювач, що регулюється; АД – амплітудний детектор; ППС – підсилювач постійного струму; ФНЧ – фільтр нижніх частот; $U_{ВХ}$ – напруга вхідного сигналу; $U_{ВИХ}$ – вихідний сигнал; $E_З$ – напруга затримки детектування; $E_Р$ – напруга регулювання.

Використання АРП зі зворотнім зв’язком дає великий динамічний діапазон, малий рівень струму регулювання та малу залежність другорядних

параметрів підсилювача від напруги регулювання. Але за допомогою зворотного регулювання можна отримати лише наближену до ідеальної характеристику АРП, тому що для роботи такого типу конструкції вкрай необхідний приріст вихідної напруги. Також така система є інерційною, тобто їй необхідний деякий час відновлення усталеного стану на виході після стрибка напруги сигналу.

У підсумку можна зауважити, що для отримання широкого динамічного діапазону та захисту приймача від перевантаження будемо використовувати підсилювач з комбінованою системою АРП, де зібрані переваги інших видів системи: отримання ідеальної характеристики регулювання та стабільний зворотний зв'язок. Акустичні сигнали від МЛА не є швидкозмінними, тому ефектом інерційністі АРП можна знехтувати. Зараз існує багато різноманітних варіантів реалізації таких систем, тому для проведення експерименту доцільніше використовувати готовий підсилювач у вигляді інтегральної мікросхеми, аніж розробляти окрему електрично-принципову схему.

Проаналізувавши ринок було обрано модуль на мікросхемі *MAX9814* — це високоякісний мікрофонний підсилювач із автоматичним регулюванням підсилення сигналу, вбудованим стабілізованим зсувом напруги малошумного джерела живлення мікрофона та з можливістю програмування коефіцієнта підсилення вихідного сигналу через цифровий вихід (рис. 2.14).

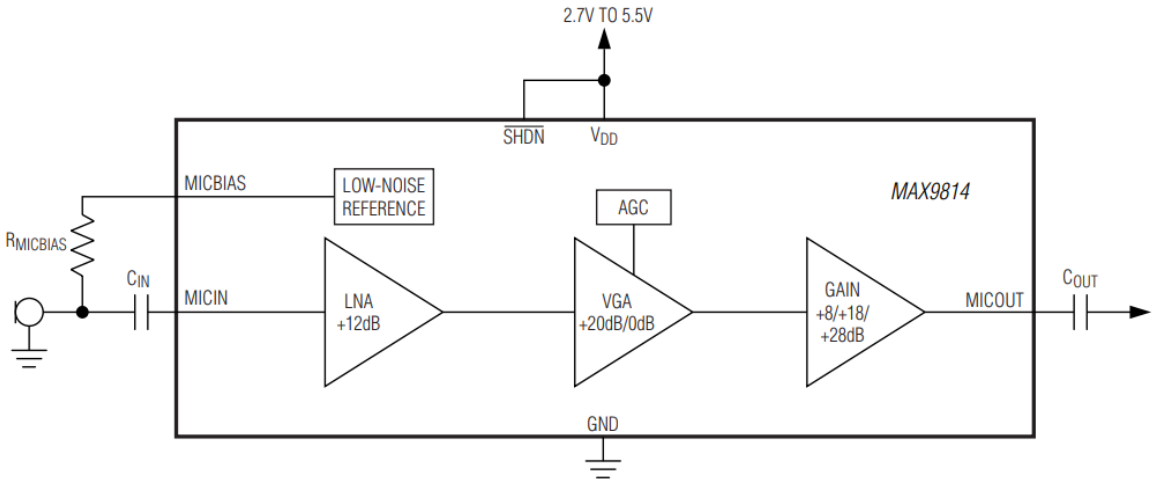


Рисунок 2.14 — Структурна схема модуля на мікросхемі MAX9814 [30]

Пристрій складається з кількох окремих схем: попереднього підсилювача з низьким рівнем шуму, підсилювачем із змінним підсиленням, вихідного

підсилювача, генератора напруги зміщення мікрофона та схеми керування АРП, для придушення сусідніх «гучних» звуків, щоб вони не давили корисні сигнали і не перевантажували підсилювач, тому цей підсилювач підходить для запису або виявленню звука, що змінюється у широких межах.

Внутрішній генератор напруги зміщення мікрофона забезпечує зміщення 2 В, яке підходить для більшості електретних та конденсаторних мікрофонів. MAX9814 підсилює вхідний сигнал у три етапи. На першому етапі вхідний сигнал буферизується та посилюється у малошумному передпідсилювачі із фіксованим коефіцієнтом підсилення 12 дБ. Другий етап складається з підсилювача зі змінним підсиленням (VGA), що керується АРП, які у комбінаційній взаємодії здатні регулювати підсилення від 20 дБ до 0 дБ залежно від вихідної напруги та порогу АРП. Останнім етапом є вихідний підсилювач, у якому відбувається фіксоване підсилення на 8 дБ, 18 дБ або 20 дБ, що встановлюється через один трирівневий логічний вхід. Без компресії з боку АРП, MAX9814 здатний забезпечувати підсилення 40 дБ, 50 дБ або 60 дБ. Поріг АРП задається зовнішнім резистивним дільником напруги. Трирівневий цифровий вхід програмує співвідношення швидкості спрацьовування до швидкості відновлення автоматичного регулятора підсилення, встановленні значення по замовчуванню від 1: 4000 до 1: 2000 або 1:500. Час затримки АРП фіксований і дорівнює 30 мс.

Основні характеристики [30]:

- Напруга живлення: 2,7 В – 5,5 В (при струмі 3 мА);
- Діапазон робочих частот: 20 Гц – 20 кГц;
- Співвідношення сигнал/шум: 61 дБ;
- Динамічний діапазон: 60 дБ;
- Вбудоване малошумляче джерело зміщення мікрофона з напругою 2 В.
- Малі загальні гармонічні спотворення: 0,04% ;
- Низька вхідна щільність шуму 30 $нВ/\sqrt{Гц}$;
- Діапазон робочих температур: -40°C … +85°C;

- Автоматичне регулювання підсилення, три коефіцієнта підсилення: 40дБ, 50дБ або 60дБ.

Вихідна напруга підсилювача становить близько 2 В при зміщенні 1,25 В, тому його можна легко використовувати з будь-яким аналого-цифровим перетворювачем із входом до 3,3 В. Для підключення АЦП до лінійного входу — просто необхідно послідовно використовувати розділовий конденсатор 100 мкФ.

2.3. АЦП. Дискретизація сигналу

Аналогові акустичні сигнали від МЛА, що були зареєстровані із використанням мікрофонів, для подальшої обробки на обчислювальних пристроях, спочатку необхідно перевести у цифровий вид за допомогою АЦП (рис. 2.15), а далі подавати на вхід звукової карти ПК.

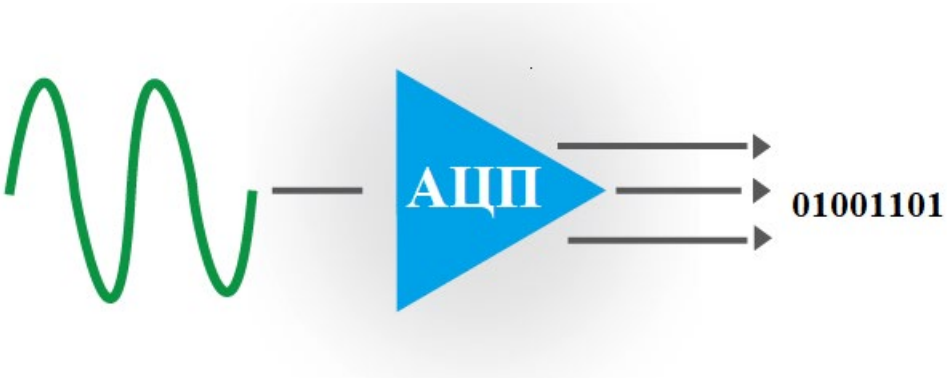


Рисунок 2.15 — Процес дискретизації сигналу за допомогою АЦП

Алгоритм аналого-цифрового перетворення наступний:

1) *дискретизація аналогового сигналу за часом* — процес заміни вихідної аналогової функції *F(t)* деякою дискретною функцією *U(nT),* тобто формується вибірка значень вихідної аналогової величини у деякі наперед задані дискретні моменти часу;

Процес дискретизації аналогового сигналу тривалістю $t_{вх}$ виконується за теоремою Котельникова, яка визначає необхідний крок дискретизації:

$$\Delta t \leq \frac{1}{2 \cdot f_m} \tag{2.1}$$

де f_m – максимальна частота спектра вхідного сигналу.

Загальна кількість кроків (M) дорівнює $M = t_{вх}/\Delta t$.

З 2.1 випливає, що аналоговий сигнал $x(t)$ з обмеженим спектром можна відновити за власними дискретними відліками, взятими з частотою більшою за подвоєну максимальну частоту спектра: $f_д > 2f_m$. Похибка перетворення відсутня на етапі дискретизації сигналу за часом.

2) *квантування за рівнем* — округлення кожного отриманого у дискретні моменти часу значення вихідної аналогової величини до деяких відомих величин – рівнів квантування N;

Процес квантування за рівнем призводить до виникнення помилки (шуму) квантування, максимальне значення якої $\pm 1/2\Delta u$ визначається розрядністю вихідного коду. При збільшенні розрядності АЦП помилка квантування може бути зменшена до скільки завгодно малої величини, але не може бути зведена до нуля вибором параметрів пристрою.

Не можна збільшувати розрядність АЦП до нескінченності, через наявність різних видів шуму. Тобто існує межа, яка обмежує мінімальну ціну поділу шкали. Іншими словами, зменшуючи розподіл шкали ми рано чи пізно «упремося» у шум. Так, звичайно, можна зробити хоч 100-бітовий АЦП, проте більшість біт даного АЦП не нестимуть корисної інформації. Саме тому характеристика роботи АЦП (рис. 2.16) має ступінчасту форму, що рівносильне наявності кінцевої розрядності АЦП.

3) *кодування* – заміна знайдених квантованих $N + 1$ значень вхідного сигналу деякими цифровими кодами.

На рисунку 2.16 аналоговий сигнал зображений як плавна лінія, а цифровий — як набір «сходинок», наближений до цієї лінії. В ідеальному випадку на графіку була б пряма лінія, тобто кожен аналоговий рівень сигналу мав би єдиний цифровий еквівалент. Чим вище частота дискретизації — тим більше «сходинок» припадає на певну ділянку плавної лінії і тим точніше цифровий сигнал відповідає вихідному аналогу.

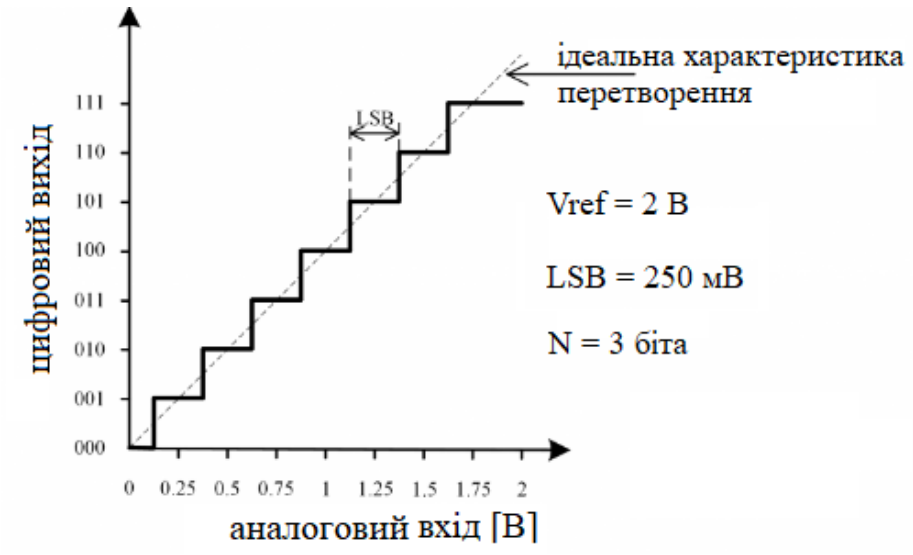


Рисунок 2.16 — Передаточна характеристика АЦП

Таким чином, високі значення даного параметра говорять про високу якість мікрофону. Однак тут потрібно сказати, що для нормального відновлення цифрового вихідного сигналу (для нормального відтворення звуку, сприйнятого мікрофоном) достатньою є частота дискретизації, що вдвічі перевищує максимальну частоту прийнятого звуку.

Варто зауважити, що підвищення частоти дискретизації позначається на обсязі переданих даних, тому високі показники не завжди є оптимальними. Проектуючи систему необхідно обирати АЦП, який би забезпечив відсутність втрати інформації під час оцифрування. Для цього розглянемо основні параметри, які впливають на вибір підходящого перетворювача. Їх можна розділити на 2 групи:

- ***Статичні*** — характеризують АЦП при постійному вхідному сигналі або такому, що змінюється дуже повільно.

- ***Динамічні*** — визначають максимальну швидкість перетворення, граничну частоту вхідного сигналу, шуми та нелінійності.

<u>Статичні параметри:</u>

- *Максимальний* ($\mathbf{V_{ref}}$) та *мінімальний* (найчастіше 0) рівень вхідного сигналу — межі діапазону шкали перетворення, по відношенню до якої оцінюється рівень вхідного сигналу (рис. 2.16);

- *Розрядність* (***n***) вихідного коду — кількість дискретних значень, які перетворювач може видати на виході АЦП після перетворення аналогового сигналу. При використанні двійкового коду розрядність розраховується за наступною формулою:

$$n = \log_2(N + 1) \tag{2.2}$$

де *N+1* — максимальна кількість кодових комбінацій (рівнів квантування) на виході АЦП;

- *Струм споживання* (***I**$_{dd}$*) залежить від частоти перетворення;

- *Молодший значущий розряд (***LSB** – *Least Significant Bit)* — мінімальний рівень вхідної напруги, яку може розрізнити АЦП. Його можна визначити за наступним виразом:

$$LSB = \frac{V_{ref}}{2^n} \tag{2.3}$$

де V_{ref} — максимальний рівень вхідного сигналу;

n — розрядність вихідного коду.

- *Помилка зміщення* (*offset error*) — відхилення передаточної характеристики реального АЦП від ідеального АЦП у початковій точці шкали. Вимірюється у долях *LSB*. При додатній помилці зміщення перехід вихідного коду з 0 в 1 виконується при вхідній напрузі відмінній від 0.5 LSB (рис. 2.17).

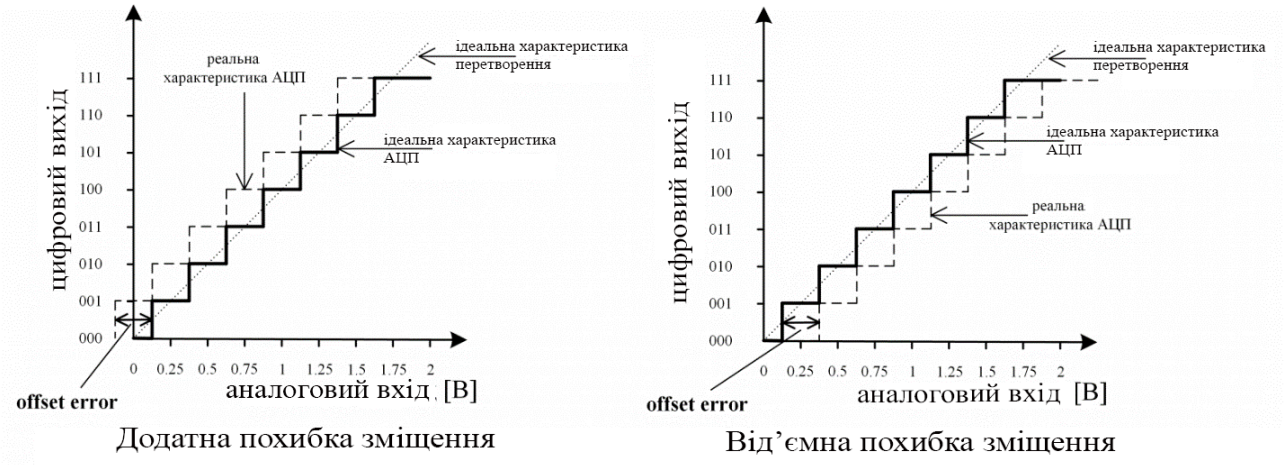


Рисунок 2.17 — Помилки зміщення

При від'ємній помилці зміщення перехід відбувається при цілих значення LSB. Обидва квантувача рівноправні, тому для простоти далі будемо розглядати лише перший варіант.

- *Помилка підсилення* (*gain error*) — відхилення середньої точки останнього кроку перетворення (якому відповідає вхідна напруга V_{ref}) реального АЦП від середньої точки останнього кроку перетворення ідеального АЦП після компенсації помилки зміщення (рис. 2.18).

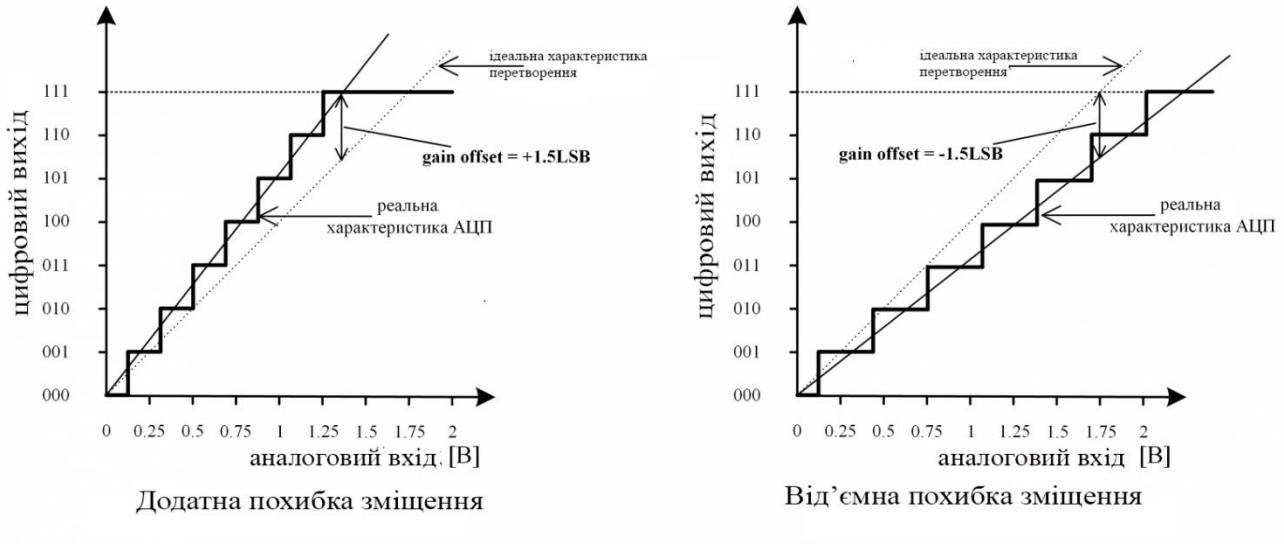


Рисунок 2.18 — Помилки підсилення

- *Загальна нескоригована помилка* (***TUE*** *– Total Unadjusted Error*) — абсолютна помилка, що включає помилки квантування, зміщення, підсилення та нелінійності. Іншими словами, це максимальне відхилення між реальною та ідеальною характеристикою перетворення. Для ідеального АЦП $TUE = 0.5LSB$ обумовлена помилкою квантування, яка виникає через округлення значення аналогового сигналу, що відповідає цифровому коду. Помилки підсилення та зміщення зазвичай роблять найбільш вагомий внесок в абсолютну помилку. Проте з погляду динамічних параметрів помилки зміщення та підсилення несуттєві, оскільки вони не породжують нелінійні спотворення.

<u>Динамічні параметри:</u>

- *Частота дискретизації* ($f_д$) — частота, на якій відбувається перетворення в АЦП. Вимірюється числом вибірок за секунду і визначається таким чином:

$$f_д = \frac{1}{T_д} \tag{2.4}$$

де $T_д$ – період вибірки.

Зазвичай під цим позначенням мають на увазі максимальну частоту дискретизації, при якій усі параметри перетворювача специфіковані (рис. 2.19).

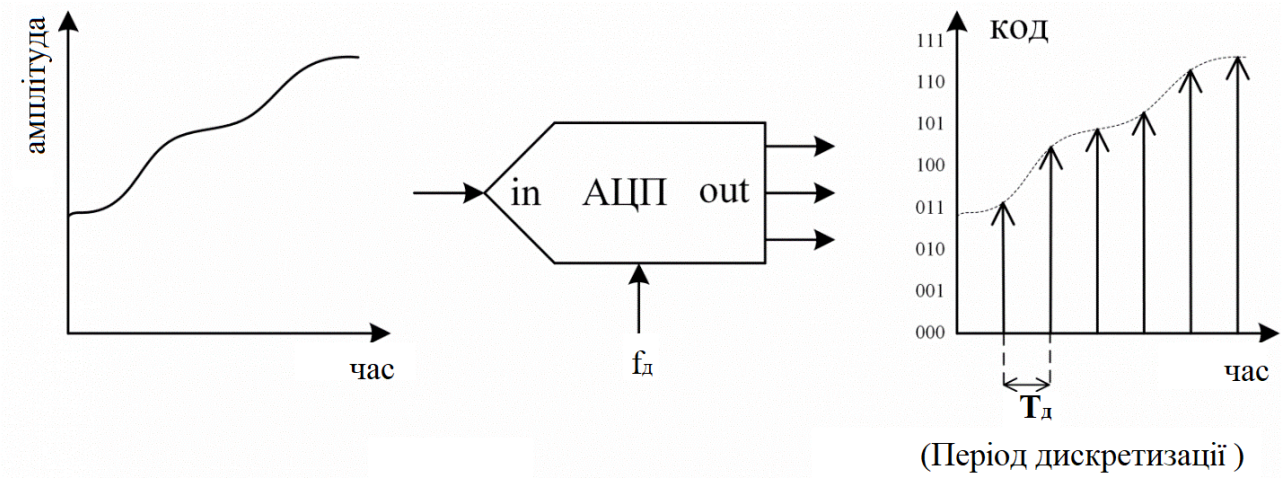


Рисунок 2.19 — Процес перетворення АЦП

- *Співвідношення сигнал/шум* (***SNR*** – *Signal-to-Noise Ratio*) визначається як відношення потужності сигналу, що обробляється АЦП, до потужності шуму, який додається в процесі перетворення (рис. 2.20). SNR зазвичай виражається в децибелах (*дБ*) і розраховується за наступною формулою: $SNR = 10 \cdot lg \frac{P_{\text{сигнал}}}{P_{\text{шум}}}$

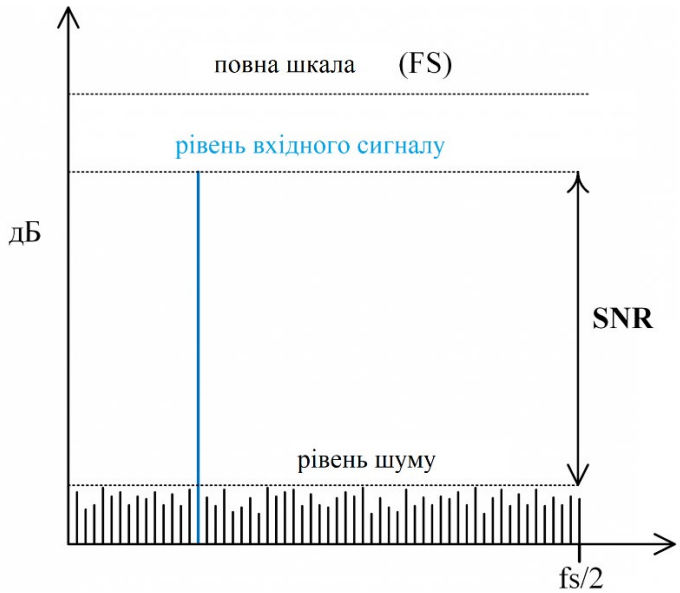


Рисунок 2.20 — Співвідношення сигнал / шум

Для оцінки SNR АЦП при розробці системи, можна скористатись наступною формулою: $SNR = 6.02N + 1.76 + 10 \cdot lg \frac{f_{\text{д}}}{2BW}$.

Перші 2 доданки враховують рівень сигналу та помилку квантування (формула справедлива для сигналу з розмахом повної шкали). Третій доданок

враховує ефект передискретизації (виграш по обробці або *processing gain*): якщо смуга оброблюваного сигналу ($BW < f_д/2$), то, застосувавши цифровий фільтр низьких частот (або смуговий, тут все залежить від смуги та несучої) до результату перетворення, можна вирізати частину шуму АЦП, а частина, що залишилася, буде розподілена від 0 до BW (рис. 2.21). Якщо шум АЦП рівномірно розподілений по всіх частотах (так званий «білий» шум), інтегральний шум після фільтрації зменшиться у $f_д/2BW$ разів, що і відображає третій член формули.

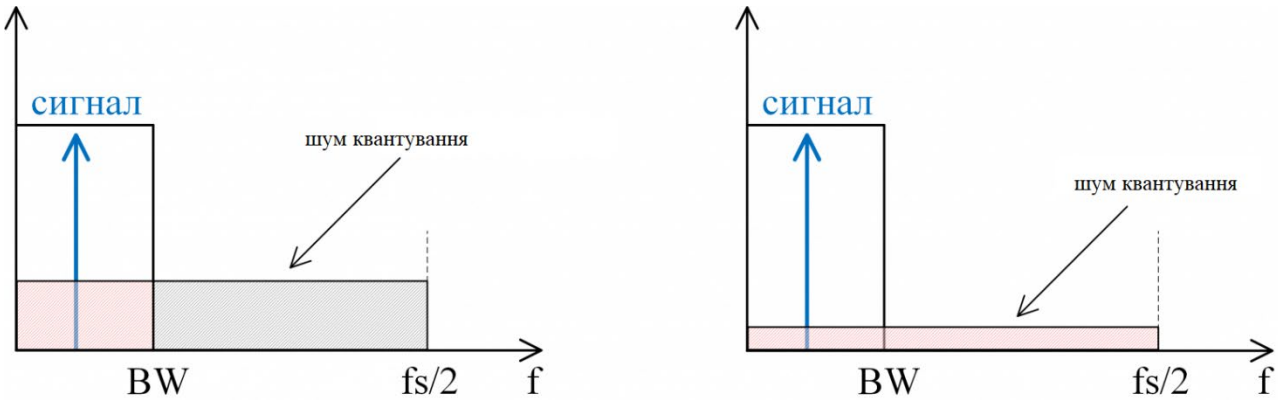


Рисунок 2.21 — Збільшення SNR за рахунок передискретизації

- *Смуга пропускання перетворювача* (***FPBW*** *– Full Power (Analog) Bandwidth*). Зазвичай ширина смуги перетворювача становить кілька зон Найквіста. Цей параметр має бути у специфікації, але, якщо його немає, можна спробувати самостійно оцінити мінімально можливе значення смуги пропускання для АЦП. За період вибірки ємність ПВЗ (пристрою вибірки та зберігання сигналу) повинна зарядитися з точністю 1·LSB. Якщо період вибірки дорівнює $1/(2f_д)$, то помилка вибірки сигналу повної шкали дорівнює: $V_{FS} \cdot e^{\left(-\frac{t}{\tau}\right)} = 1 \cdot LSB$. Розв'язавши відносно t, отримаємо: $t = -\tau \cdot ln\left(\frac{1 \cdot LSB}{V_{FS}}\right)$. Поклавши, що $\tau = \frac{1}{2\pi \cdot FPBW}$, визначимо мінімальну смугу АЦП (для $t = 1/(2f_s)$):

$$FPBW = -\left(\frac{f_s}{\pi}\right) \cdot ln\left(\frac{1 \cdot LSB}{V_{FS}}\right)$$

На даний момент існує багато різних архітектур АЦП, але немає такої, яка досягала максимальних значень всіх, описаних вище параметрів. У таблицю 2.2 занесено переваги та недоліки кожної архітектури перетворювача.

Таблиця 2.2 — Переваги та недоліки основних архітектур АЦП

Вид АЦП	
flash	*Переваги:* - швидкий перетворювач; - перетворення здійснюється в один такт.
	Недоліки: - високе енергоспоживання; - обмежений дозвіл; - потребує великої площі кристала ($2^N–1$ компараторів); - важко узгодити багато елементів (як наслідок низький вихід придатних).
folding-interpolated	*Переваги:* - швидкий перетворювач; - перетворення здійснюється в один такт; - потребує меншу кількість компараторів завдяки попередній «згортці» всього діапазону обробки в деякий вужчий діапазон; - займає менше площі.
	Недоліки: - помилки, пов'язані з нелінійністю блоку згортки. - затримка на встановлення рівнів у блоці згортки, яка зменшує максимальну $f_д$. - середня роздільна здатність.
SAR	*Переваги:* - висока точність; - низьке енергоспоживання; - легке у використанні.
	Недоліки: - обмежена швидкість.
Pipeline	*Переваги:* - швидкий перетворювач; - найвища точність серед швидких АЦП; - чи не займає велику площу; - має менше споживання, серед аналогічних швидких перетворювачів.
	Недоліки: - конвеєрна затримка.
Dual-slope	*Переваги:* - середня точність перетворення; - простота конструкції; - низьке споживання; - стійкість до змін факторів довкілля.
	Недоліки: - обробляє низькочастотні; - сигнали (низька $f_д$); - посередній дозвіл.
∑-Δ	*Переваги:* - Найвища точність перетворення завдяки ефекту "Noise shaping" (специфічна фільтрація шуму квантування) та передискретизації.
	Недоліки: - Не може працювати з широкосмуговим сигналом.

Для вибору підходящої моделі АЦП, визначимо його основні параметри.

Частота дискретизації обчислюється за виразом 2.4, де період вибірки зворотно пропорційний подвоєній максимальній робочій частоті мікрофону ($f_{\text{роб}}$) згідно 2.1. Оскільки за технічним паспортом обрана модель мікрофону працює у діапазоні до 20 кГц, то частота дискретизації: $f_{\text{д}} = f_{\text{роб}} \cdot 2 = 40$ кГц.

Шумові характеристики АЦП відносяться до кількості небажаного шуму у вихідних даних. АЦП з низьким рівнем шуму є кращим для застосувань, де потрібна висока точність. *ЕРС шуму* ($\varepsilon_{ш}$) АЦП визначимо за наступним виразом:

$$\varepsilon_{\text{ш}} = e_{\text{ш}} \cdot \sqrt{f_{\text{роб}}} \cdot k \qquad (2.5)$$

Оскільки робоча смуга частот 20 кГц, а, згідно до даташиту [30], вхідний рівень шумів мікросхеми на цій частоті $e_{\text{ш}} = 30\ \frac{\text{нВ}}{\sqrt{\text{Гц}}}$, то при мінімальному коефіцієнті підсилення (k = 40 дБ = 100 разів) ЕРС шуму дорівнює 0,43 мВ.

Максимальний розмах сигналу, що забезпечує мікросхема 2 В (-1 В та 1 В — мінімальне та максимальне пікове значення), тому *опорна напруга* АЦП повинна дорівнювати 2,54 В, як найближче стандартне значення. Відповідно *мінімальний рівень вхідного сигналу* буде 0 В, а *максимальний* ($\boldsymbol{V_{ref}}$) — 2,54 В.

Кількість рівнів квантування (***N***) визначається як значення опорної напруги АЦП поділене на ширину сходинки, тобто величину напруги молодшого біту (2.3), або на значення ЕРС шуму АЦП (2.5) :

$$N = \frac{V_{ref}}{LSB} = \frac{V_{ref}}{\varepsilon_{\text{ш}}} = 5893$$

Розрядність (***n***) вихідного коду АЦП за формулою 2.2 дорівнює 12,25. Відповідно необхідно брати АЦП з розрядністю не менше 12, тоді шум квантування вже суттєво впливати не буде, але існує ймовірність втрати частини інформації через меншу кількість рівнів квантування. Для уникнення такої ситуації, краще використовувати 14- або 16-розрядне АЦП, які на сьогодення є найбільш прийнятними для перетворення звукових сигналів. Використання АЦП з більшою розрядністю не є доцільним, оскільки через імпульсні завади або теплові шуми молодші біти не будуть нести корисну інформацію.

Для отримання звуку з низьким рівнем шумів та спотворень під час запису бажано використовувати зовнішню звукову карту. При побудові установки для експериментів, аудіосигнал після підсилення з мікросхеми *MAX9814* поступає на лінійний вхід 6-ти канальної зовнішньої звукової карти моделі *CM6206*, яка через інтерфейс USB під'єднана до ПК, де відбувається процедура обробки прийнятих сигналів. *Інтерфейс* АЦП у звуковій карті використовує стандарт S/PDIF, що підтримує передавання цифрових аудіосигналів від одного пристрою до іншого без процедури перетворення в аналоговий сигнал, що дає змогу уникнути погіршення якості звуку і є цілком сумісним з рештою системи. Перевагою типу S/PDIF є відмінна стійкість до електричних перешкод.

Основні технічні характеристики зовнішньої звукової карти CM6206 [31]:

- Апаратний декодер звуку;
- Підтримка: 5.1 звук;
- Інтерфейс під'єднання: USB 2.0;
- Енергоспоживання: живлення від USB порту;
- ЦАП/АЦП: розрядність 16-біт та частота дискретизації 48 кГц (що цілком задовольняє висунутим вище вимогам)

2.4. Формування бази сигналів МЛА

Запис аудіосигналів польотів малих безпілотних літальних апаратів, що зареєстровані акустичним детектором, було зроблено на ПК із використання програмного забезпечення (ПЗ) «*Adobe Audition*». Інтерфейс цього ПЗ також дозволяє легко зберігати записи у будь-якому місці на ПК, відкривати та програвати записані раніше сигнали, а також дивитись їхні спектрограми.

2.4.1. Опис вимірювальної установки

Для реєстрації звукових сигналів від МЛА на відкритому повітрі на висоті 1 м над землею був встановлений всеспрямований електретний мікрофон із поролоновою вітрозахисною насадкою. Він підключався до мікрофонного модуля MAX9814 та, через зовнішню звукову карту CM6206, передавав отриманий сигнал на ПК, який був налаштований на запис аудіо файлів з

частотою дискретизації 48 кГц. Мікрофон та модуль живляться від акумулятора з напругою до 5 В, звукова карта через інтерфейс USB — від комп'ютера, який в свою чергу може живитися від батареї або від місцевих розеток. Для створення набору аудіоданих звуків МЛА, були використані декілька моделей квадрокоптерів: «*DJI Mavic 2 Pro*» та «*DJI Mavic 3*», які зараз активно використовується військовими для розвідки, через свої малі габаритні розміри та дизайн корпусу з чотирма двигунами, що у поєднанні з добре збалансованими легкими пропелерами зменшує рівень шуму під час польоту до цілі, дрон-камікадзе *«FPV»* та цивільні коптери *«FeiLun FX137»* і «*LH X43-W*».

2.4.2. Умови проведення експерименту

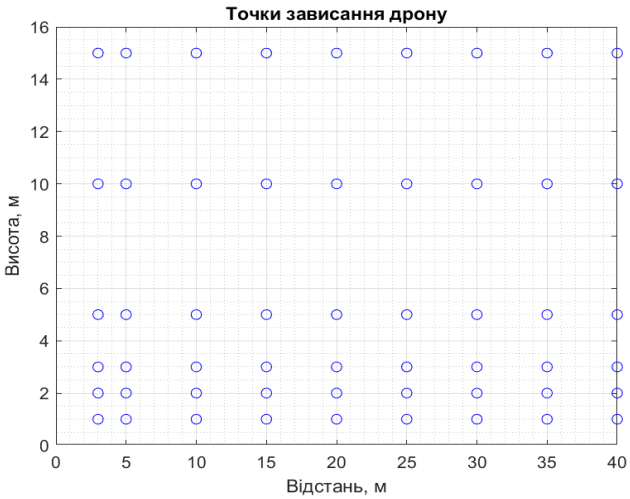


Рисунок 2.22 — Координати позицій зависання МЛА

Було проведено випробування у статичному та динамічному режимах. У статичному режимі квадрокоптери зависали на 1 хвилину на заздалегідь визначених координатних точках, що змінюються за відстанню та висотою (рис. 2.22). Записи були позначені висотами та відстанями у 2D площині, з мікрофоном на початку координатної площини. У динамічному режимі, дрон рухався прямо назустріч мікрофону з дистанції 250 м, але на одній висоті 5 м.

Тестові польоти здійснювались на відкритій площадці полігону за містом, без промислових об'єктів поряд, в теплу погоду з температурою повітря 25 ºС. Під час запису відчувались невеличкі шуми поривів вітру, комашок та руху транспорту по трасі, що розташована на відстані 1 км.

2.4.3. Результати експерименту

У результаті експерименту було згенеровано 1866 записів зависання дрону «*DJI Mavic 3*», 1408 моделі «*DJI Mavic 2 Pro*», 1047 — «*FPV*», 83 — *«FeiLun FX137»* і 56 — «*LH X43-W*». Довжина кожного фрагменту 1 секунда з площею перекриття 50%.

Оброблення отриманих аудіозаписів відбувалось із використанням бібліотек програмного забезпечення *MATLAB*. На рисунку 2.23 наведені приклади результуючих графіків спектрів щільності потужності сегментів аудіосигналів квадрокоптера «*DJI Mavic 3*» розрахованих методом Велча на деяких відстанях до мікрофону. Для мінімізації ефекту розповсюдження звуку та надання вектору-об'єкту значного інваріантного характеру відносно відстані, кожна компонента вектору нормалізується відносно суми величин двох найвищих гармонік у множині.

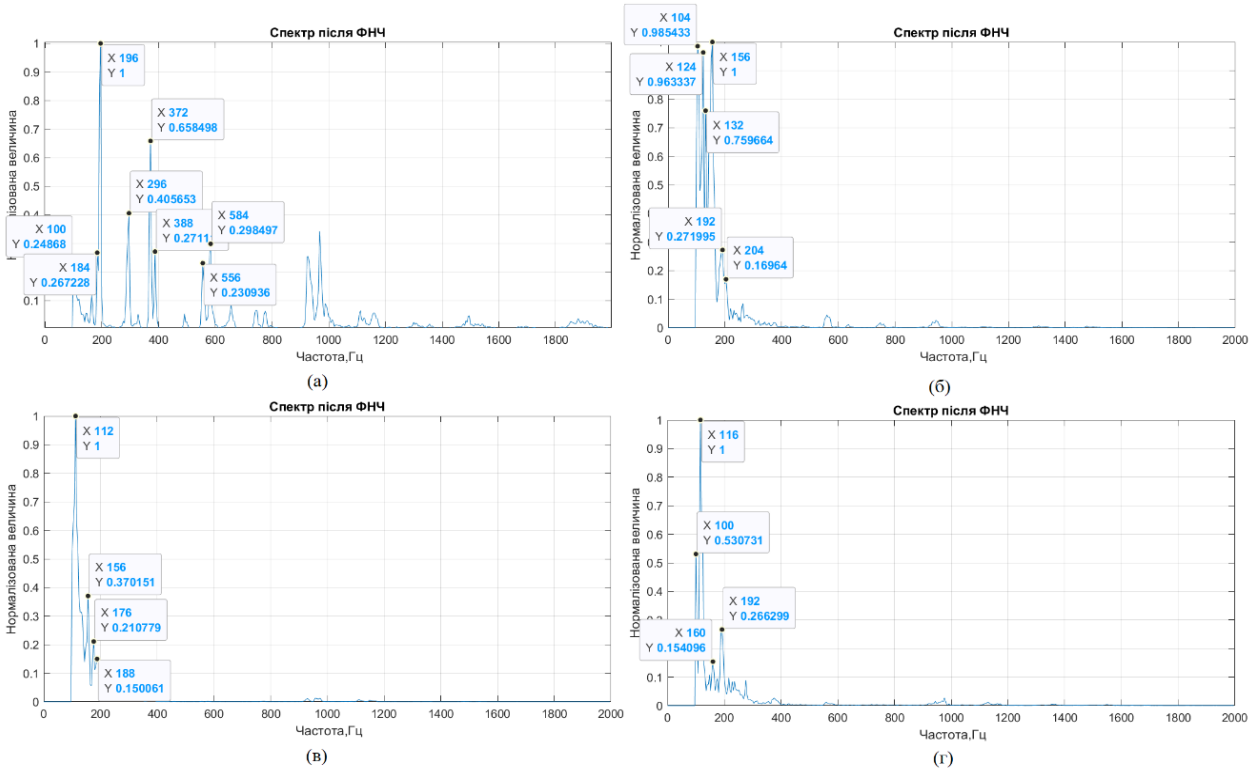


Рисунок 2.23 — Спектр акустичного випромінювання квадрокоптера *«DJI Mavic 3»* на відстані від мікрофону: а) 10 м, б) 20 м, в) 30 м, г) 40 м

З отриманих графіків акустичного випромінювання квадрокоптера (рис. 2.23) підтверджується твердження, що МЛА в режимі польоту мають періодичні спектральні компоненти, а кількість двигунів відповідає кількості основних тонів у випромінюванні, що є одними із найголовніших їх ознак, на яких базуються методи ідентифікації та класифікації дронів із використанням енергетичного підходу [32]. Кожний гармонічний тон випромінювання дрону f_i визначається як добуток частоти обертання електричного мотора (основної

гармоніки) f_0, порядкового номера відповідної гармоніки i та числа лопатей пропелера N: $f_i = N \cdot f_0 \cdot i$.

Періодичні спектральні максимуми дрону складаються з чотирьох основних тонів та періодичних груп гармонічних складових, що розташовані близько один до одного та з часом змінюють своє положення на частотній осі, у зв'язку з вітровим навантаженням на пропелери. Воно впливає на швидкість обертання роблячи її різною для кожного з електродвигунів. Також з рисунку 2.23 видно, що незалежно від відстані, основна інформація про об'єкт лежить у діапазоні частот від 100 Гц до 1500 Гц, це підтверджує експеримент у роботі [33]. Спектральні компоненти шуму вітру знаходяться на частотах до 250 Гц [34], тому для його компенсування варто використовувати захисну насадку на мікрофон, щоб збільшити дальність та ймовірність виявлення МЛА.

Варто зауважити, що зі збільшенням відстані від мікрофона до об'єкту, значно зменшується відносний рівень гармонічних складових тонів випромінювання дронів, а також відбувається ефект порушення гармонічними частотами принципу кратності основним тонам. На відстанях від 30 м, деякі складові основних тонів випромінювання, а тим паче їхні гармоніки, зникають і завдання виявлення дрону гармонічним способом сильно ускладняється. Саме тому надійніше базувати алгоритми виявлення безпілотників на аналізі енергетичної характеристики, яка зважає на вплив усіх складових спектру випромінювання, а також незалежна від виду МЛА.

Основним критерієм оцінки роботи акустичного детектора є якість отриманих сигналів від МЛА на максимальній дальності. На відстані до 40 м дані підтверджують присутність частот основного тону випромінювання, кількість яких збігається з кількістю електромоторів дрону. Ці характеристики є одними із найголовніших ознак, на яких варто основувати методи ідентифікації та класифікації дронів із використанням енергетичного підходу.

Варто враховувати, що зараз на дистанції до об'єкту більшій, ніж 40 м, у спектрі сигналу будуть наявні не всі головні тони випромінювання, наприклад, лише 3, у цьому випадку квадрокоптер буде класифікований, як коптер з трьома

пропелерами. Незважаючи на це з упевненістю можна говорити про наявність ворожого МЛА у просторі сканування, отже із завданням виявлення об’єктів розроблювальний детектор справляється.

Отримана під час експерименту бібліотека аудіофайлів буде застосовуватись у якості тренувальної вибірки для навчання штучної нейронної мережі, щоб збільшити максимальну дальність детектування та ефективність класифікації безпілотників.

Висновки до розділу 2

1. У другому розділі дисертації була створена структурна схема акустичного детектора.

2. У якості приймального пристрою для реєстрації звукових хвиль від малих безпілотних літальних апаратів було обрано електретний мікрофон моделі ***CMA-4544PF-W***.

3. Обґрунтовано у якості підсилювача з АРП використати мікросхему *MAX9814.*

4. Було визначено основні параметри АЦП для якісного перетворення зареєстрованих акустичних сигналів: мінімальну частоту дискретизації – 40 кГц та розрядність вихідного коду не менше 12 біт.

5. Запропоновано використовувати зовнішню звукову карту ***CM6206*** із лінійним входом.

6. У підсумку польових випробувань була створена база аудіофайлів шуму квадрокоптерів «*DJI Mavic 3*», «*DJI Mavic 2 Pro*», «*FPV*», *«FeiLun FX137»* та «*LH X43-W*».

7. Основний результат розділу 2 опубліковано автором дисертації у науковій статті [35].

РОЗДІЛ 3. РОЗРОБКА АЛГОРИТМУ ДЛЯ ОБРОБЛЕННЯ АУДІОСИГНАЛІВ ІЗ ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ

У розробленні алгоритму виявлення та класифікації аудіосигналів малих безпілотних літальних апаратів із використанням методів машинного навчання є кілька основних етапів (рис.3.1):

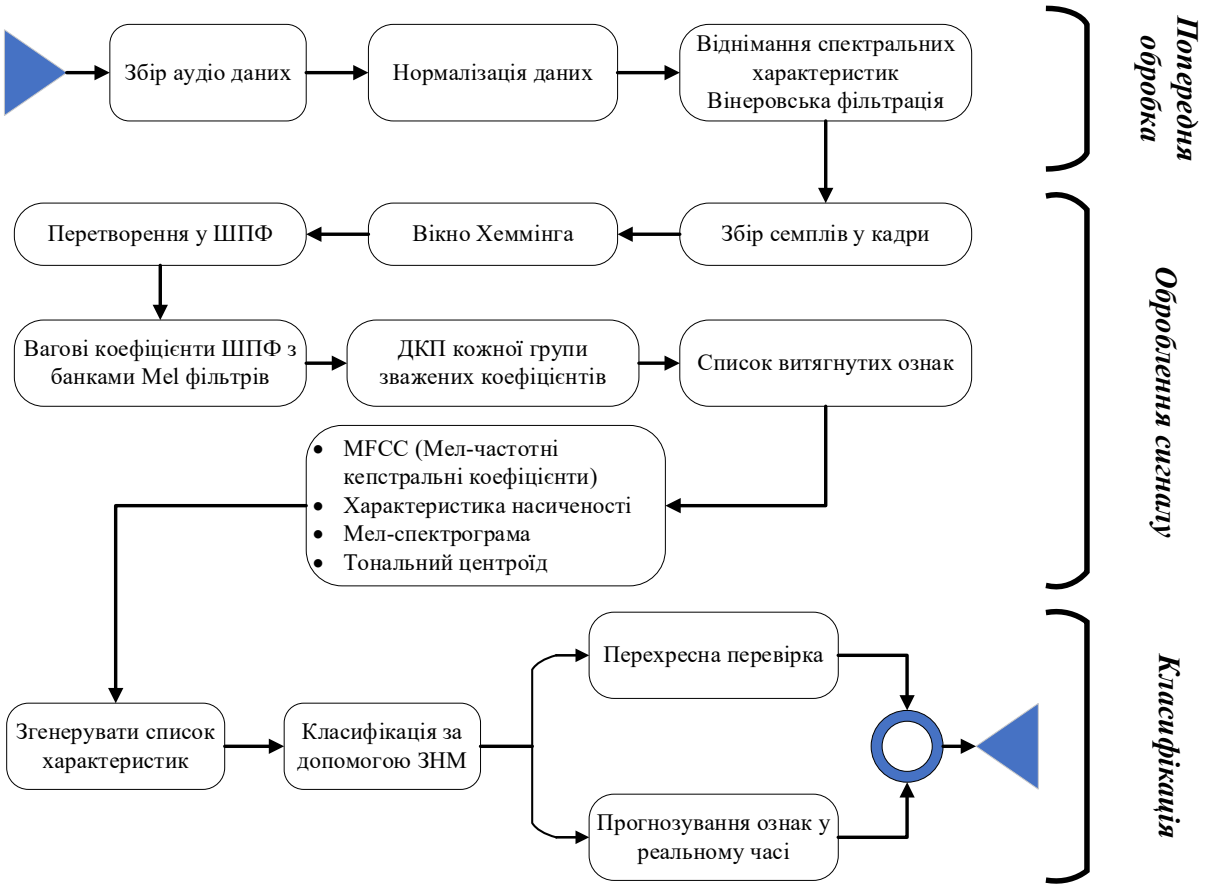


Рисунок 3.1 — Блок-схема виявлення та класифікації акустичних сигналів МЛА

1. ***Створення набору даних*** аудіозаписів сигналів цілі, а також фонових шумів та інших звуків, що можуть бути присутніми під час процесу виявлення.

2. Застосування ***методів попередньої обробки*** до аудіосигналів, щоби підвищити їхню якість для подальшого аналізу. Цей процес включає в себе передискретизацію, зменшення рівня шуму, нормалізацію і фільтрацію сигналів, а також кадрування та віконне керування.

3. ***Вилучення*** відповідних ***характеристик*** окремо з кожного кадру попередньо оброблених аудіосигналів, щоби представити аудіо данні спрощеним набором акустичних ознак. До поширених характеристик аудіо належать спектрограми, мел-спектрограма, мел-частотні кепстральні коефіцієнти, характеристика насиченості та тональні дескриптори. Ці характеристики фіксують часові та спектральні характеристики аудіосигналів.

Через присутність безшумних кадрів із кожного аудіофайлу вибирається фіксоване число найбільш енергоємних (гучних) кадрів, відкидаючи інші. Кількість кадрів, що надходять із кожного файлу, незалежно від його довжини, за такого підходу залишається незмінною. Для моделювання динамічних властивостей звуків також мають враховуватися суміжні кадри. Кількість ознак, які становить кожен кадр, збільшується через поєднання ознак поточного кадру з певною кількістю сусідніх (лівого та правого).

4. ***Підготовка навчальних даних***. Відбувається розподіл даних на три відокремлені набори: для навчання, перевірки (валідації) та тестування [36].

5. ***Створення та навчання моделі*** машинного або глибокого навчання, використовуючи навчальні дані та витягнуті характеристики аудіосигналу.

6. ***Оцінювання продуктивності*** навченої моделі, використовуючи валідаційний набір даних, та ***оптимізація моделі*** за допомогою налаштування гіперпараметрів для підвищення точності і здатності до узагальнення. Цей крок включає коригування архітектури мережі, швидкості навчання, методів регуляризації або інших специфічних для моделі параметрів.

7. ***Тестування*** — оцінювання продуктивності та здатності до узагальнення фінальної навченої моделі на тестовому наборі даних. Цей крок дає змогу оцінити, наскільки добре модель працюватиме за нових, раніше не бачених даних, взятих із того ж самого розподілу, що й навчальні дані.

8. Застосування методів ***пост-обробки*** для уточнення результатів класифікації: фільтрацію помилкових спрацьовувань, застосування порогових значень для прийняття рішень або використання інших методів для підвищення загальної точності класифікації.

9. ***Деплоймент*** навченої моделі для обробки аудіосигналів у режимі реального часу або пакетної обробки, залежно від конкретного застосування. Розгортання ПЗ готової моделі може включати інтеграцію алгоритму в більшу систему або використання його, як окремого інструменту.

Проаналізуємо більш детально кожний етап алгоритму розроблення програмного забезпечення для оброблення акустичних сигналів малих безпілотних літальних апаратів.

3.1. Створення набору аудіоданих моделей МЛА

Створення набору та підготовка аудіоданих для подальшої роботи зазвичай відбувається в наступній послідовності:

1. Висування вимог до даних

У якості *цільових сигналів* використовуються аудіозаписи шумів польоту та зависання в просторі різних моделей МЛА (розділ 2, п. 2.4.1), а також фонові шуми навколишнього середовища: вітру, автомобілів, людей тощо.

Нейронні мережі (НМ), з високою продуктивністю моделі, особливо залежать від доступності великої кількості навчальних даних для вивчення нелінійної функції від входу до виходу, яка добре узагальнює і дає високу точність класифікації за прихованими об'єктами. Тож, чим більше довжина кожного аудіофайла та їхня кількість, тим більша ефективність роботи алгоритму. Формат файлів має бути *WAV* або *MP3* через широкий діапазон бітрейта, що підтверджує високу якість записів.

2. Вибір обладнання для запису

Вибір відповідного обладнання для запису, випливає з вимог поставлених до даних та бюджету. Це можуть бути, як мікрофони та аудіоінтерфейси професійного рівня, так і більш доступні пристрої споживчого класу або навіть додатки для смартфонів із функцією аудіозапису. Створення власного пристрою для запису аудіо сигналів розглянуто в *розділі 2*.

3. Налаштування середовища для запису

У якості середовища для запису аудіосигналів використовувалися програми *Matlab* та *Adobe Audition*, які надають можливість записувати файли в необхідних форматах. Фактори впливу зовнішніх джерел шуму довкілля, план проведення експерименту, часові рамки та умови збору даних детально розглянуто в *пункті 2.4.*

4. *Організація та позначення даних*

Після запису аудіозразки необхідно структурувати, а їхні колекції зберегти в базу аудіосигналів (*AudioDatastore*). Чітка структуризація та точне маркування даних має важливе значення для подальшого навчання та оцінювання моделей машинного навчання. AudioDatastore завжди є папкою даних верхнього рівня на ПК та містить підпапки, кожна з яких названа на честь відповідної моделі МЛА або фонового шуму, яку вона представляє, і має складатися лише з колекції аудіофайлів, що відповідають цій назві. Імена аудіофайлів, які містять корисну інформацію про об’єкт, створюються з урахуванням назви моделі МЛА або виду фонового шуму, класу, умов запису та будь-яких інших відповідних метаданих.

5. *Розширення даних*

Щоб збільшити розміри та різноманітність набору даних, потрібно використати метод аугментації — застосування однієї або декількох деформацій до колекції анотованих навчальних даних, які призводять до створення нових або додаткових матеріалів. Ключова концепція аугментації в тому, що деформації, які застосовуються до позначених даних, наприклад, зсув висоти тону, розтягування в часі або додавання штучного шуму, не змінюють семантичного значення міток. Якщо навчати НМ додатковими спотвореними даними, то вона стане нечутливою до цих деформацій і зможе краще узагальнити приховані дані.

3.2. Попередня обробка вхідних аудіосигналів МЛА

Аудіосигнали, що надходять у режимі реального часу або завантажені з файлів у форматі WAV / MP3, необхідно перетворити у вигляд, придатний для подачі на вхід НМ. Для цього використаємо методи попередньої обробки аудіо:

- *Нормалізація даних*

Щоб уникнути зміщення, спричиненого варіаціями амплітуди вхідного аудіосигналу, виконаємо його нормалізацію до діапазону від -1 до 1, це забезпечить постійний рівень сигналу.

- *Фільтрація*

Для видалення небажаних частотних компонентів сторонніх шумів, покращення певних характеристик аудіосигналів або підкреслення необхідного для аналізу частотного діапазону, потрібно застосовувати високочастотні (ФВЧ), низькочастотні (ФНЧ) або смугові фільтри.

- *Зменшення шуму*

Найбільш ефективними методами для підвищення чіткості аудіосигналу, зменшення рівня фонових шумів та завад, є спектральне віднімання, фільтр Вінера та адаптивна фільтрація.

Метод *спектрального віднімання* оцінює очищений спектр аудіосигналу, через віднімання спектра, що містить лише шум, із зашумленого оригінального сигналу. Перевагами цього методу є простота та обчислювальна ефективність для стаціонарного шуму або шуму з повільно змінним спектром. З мінусів виділяють можливість внесення спотворень в оброблений сигнал, а також нижчу ефективність для нестаціонарних або швидкозмінних шумів.

Вінеровська фільтрація — це статистичний підхід, який мінімізує середньоквадратичну похибку між вихідним і відновленим сигналом. Такий підхід оцінює співвідношення сигнал/шум для кожної частотної складової і застосовує фільтр, який адаптується до оціненого значення С/Ш.

Фільтр Вінера забезпечує краще зменшення рівня шуму, порівнюючи зі спектральним відніманням, особливо для нестаціонарного шуму, а також може зберігати більше характеристик вихідного сигналу. Але використання такого фільтру зазвичай вимагає знання статистики сигналу й шуму, яка не завжди може бути доступною. Також до недоліків можна додати вищу обчислювальну складність та ресурсозатратність, ніж у спектрального віднімання.

Метод *адаптивної фільтрації* використовує адаптивні алгоритми для оцінювання шуму в реальному часі на основі вхідного аудіосигналу й адаптивно

віднімає оцінений компонент шуму. Він безперервно оновлює коефіцієнти фільтра, щоб відстежувати та адаптуватися до змін у характеристиках шуму.

Такий фільтр може адаптуватися до різних шумових умов у реальному часі, він ефективний для нестаціонарного шуму і швидко змінного середовища. Головним недоліком такого підходу є необхідність достатньої кількості навчальних даних, тому можуть виникнути проблеми зі збіжності в певних сценаріях. Також у адаптивної фільтрації вища обчислювальна складність, порівнюючи зі спектральним відніманням та фільтрацією Вінера.

На короткому інтервалі часу 15-30 мс, аудіосигнали шуму МЛА є локально стаціонарними, відповідно час та їхня спектральна поведінка вважаються практично однорідними на проміжку в декілька мілісекунд, тому методи спектрального віднімання та Вінеровської фільтрації забезпечують найкраще зменшення рівня сторонніх шумів за таких умов.

- *Видалення тиші*

Видалення або обрізання періодів тиші з аудіосигналів може бути корисним для зменшення обчислювальних ресурсів або зосередження аналізу лише на активних ділянках сигналу.

- *Попереднє виділення*

Ця техніка передбачає застосування до аудіосигналу ФВЧ, який підсилює високочастотні компоненти сигналу і зменшує енергію нижчих частот. Основними мотивами використання попереднього підсилення є вирівнювання частотної характеристики та покращення співвідношення С/Ш. Високочастотні компоненти, як правило, несуть більше інформації і мають вище співвідношення С/Ш у порівнянні з низькочастотними. Збільшуючи високочастотний контент, попереднє виділення може підсилити важливі компоненти сигналу й поліпшити їхнє виявлення та розбірливість, особливо в присутності фонового шуму. Частота зрізу фільтру має бути не менше 10 кГц.

- *Сегментація*

Збирання відліків (семплів) у кадри означає поділ аудіосигналу на менші сегменти або кадри фіксованої тривалості, зазвичай по 15-30 мілісекунд із

деяким перекриттям (наприклад, 50%), що дає змогу якісно дослідити характеристики сигналу в часі. Аналізуючи сегменти на коротких часових інтервалах, можна зафіксувати зміни, динаміку та варіації сигналу, які можуть відбуватися в межах кожного кадру.

Під тривалістю сегменту (*segment duration*) розуміють довжину аудіосигналу, яка аналізується незалежно за один раз. Іншими словами, аудіосигнал розбивається на сегменти фіксованої довжини, і кожен сегмент аналізується окремо. У програмному алгоритмі тривалість сегмента дорівнює 1 секунда. Кількість відліків у кожному сегменті (*segmentSamples*) аудіосигналу вибирається на основі бажаної частотної розділової здатності аналізу та довжини аудіосигналу. Вона фіксована й зазвичай визначає тривалість аналізу ШПФ.

Тривалість кадру (*frame duration*) — це довжина часового вікна, яке використовується для вилучення частотного вмісту кожного сегмента. Часове вікно, як правило, є віконною функцією короткої тривалості, наприклад, вікном Ханна або Хеммінга, яка множиться на сегмент аудіосигналу перед обчисленням перетворення Фур'є. Тривалість цієї віконної функції і називається тривалістю кадру. У програмному алгоритмі вона встановлена на 25 мс, тож кожен сегмент тривалістю 1 с розбивається на 40 кадрів, кожен із яких триває 25 мс.

Інтервал часу між початками сусідніх кадрів називається тривалістю стрибка (*hop duration*). Після обчислення перетворення Фур'є кадру віконна функція зсувається вперед у часі на певну величину, яка називається тривалістю стрибка, і для наступного кадру обчислюється інше перетворення Фур'є. Оптимальне значення тривалості стрибка дорівнює 10 мс, відповідно, після обчислення перетворення Фур'є для кожного кадру, віконна функція зсувається вперед на 10 мс, що призводить до перекриття 15 мс між сусідніми кадрами.

Кадри, що перекриваються, забезпечують безперервність і плавні переходи між сусідніми кадрами, зменшуючи артефакти та різкі зміни в сигналі, що аналізується. Кількість відліків, що перекриваються між сусідніми кадрами *(OverlapSamples*) обчислюється, як різниця між кількістю відліків у кожному короткочасному вікні, яке використовується для обчислення перетворення Фур'є

(*frameSamples*), та кількістю відліків, на яку зсувається короткочасне вікно між послідовними кадрами (*hopSamples*) .

- *Особливості отримання спектру*

Щоб отримати спектр аудіосигналу, що змінюється в часі, необхідно розділити сигнал на фрейми та помножити кожен із них на віконну функцію.

Правильно обрана віконна функція дасть змогу:

- *зменшити спектральний виток;*

Коли сигнал аналізується в частотній області за допомогою таких методів, як перетворення Фур'є або ШПФ, різкі початкові та кінцеві точки сегментів сигналу можуть вносити паразитні частотні компоненти. Віконні функції застосовуються для плавного звуження сигналу до нуля на краях, зменшуючи спектральні витоки та мінімізуючи артефакти під час аналізу в частотній області.

- *покращити роздільну здатність;*

Зменшуючи вплив сусідніх частотних компонентів, віконні функції допомагають відокремити енергію певного частотного компонента від сусідніх. Це призводить до кращого представлення сигналу в частотній області.

- *зменшити крайовий ефект;*

У деяких випадках у разі застосування фільтрів або інших методів обробки, сигнал може спотворюватися або демонструвати артефакти поблизу країв. Віконні функції звужують сигнал до нуля на краях, мінімізуючи ці крайові ефекти й забезпечуючи більш плавні переходи.

- *компромісне керування*.

Різні віконні функції мають різні компроміси між шириною основної смуги й рівнем бічної смуги в частотній області.

Вікна Ханна й Хеммінга є гарним вибором для аналізу аудіосигналів, і вибір між ними залежить від конкретних вимог програми. Основна відмінність між вікнами полягає у формі їхніх спектральних бічних пелюсток і ширині головних пелюсток. Якщо частотна роздільна здатність важливіша за точність у часовій області, вікно Ханна може бути кращим вибором. Але в завданні

ефективного оброблення акустичних сигналів МЛА точність у часовій області важливіша, отже, обираємо вікно Хеммінга.

Після операції згладжування вікном Хеммінга добре проявляються гармонійні складові аудіосигналу, тому за допомогою алгоритму ШПФ знайдемо амплітудний спектр та всю інформацію про фазову складову сигналу.

Вибір відповідної довжини ШПФ є важливим для досягнення хорошої спектральної роздільної здатності, водночас зберігаючи розумні обчислювальні витрати на аналіз. Довжина швидкого перетворення Фур'є мусить відповідати розміру вікна аналізу — функції, яка множиться на аудіосигнал для вилучення частотного вмісту кожного короткочасного сегмента. Оскільки у якості вікна аналізу використовується вікно Хеммінга, то оптимальною довжиною ШПФ буде 2^n, що більше або дорівнює довжині вікна, щоб запобігти спектральному витоку, коли енергія від частотної складової «витікає» в сусідні біни, а також досягти хорошої частотної роздільної здатності.

3.3. Виділення ознак з вхідних аудіосигналів МЛА

Виділення звукових ознак (*Audio Feature Extractor*) — це процес вилучення значущої та релевантної інформації з необроблених аудіосигналів для того, щоби представити аудіо у вигляді, який можна легко проаналізувати та обробити алгоритмами машинного навчання. Виділені ознаки використовуються, як вхідні дані для моделі, що дає змогу їй навчатися на аудіоданих і робити прогнози щодо нових, небачених раніше аудіосигналів.

Підхід глибокого навчання розглядає неструктуровані аудіопредставлення, процес виокремлення ознак відбувається автоматично. Основними характеристиками або представленнями, які безпосередньо використовуються в архітектурі НМ, є спектрограми, мел-спектрограми та мел-частотні кепстральні коефіцієнти (MFCC). Розглянемо їхні переваги та недоліки.

Спектрограма — це візуальне зображення спектру частот звукового сигналу, який змінюється із часом (рис. 3.2). Її отримують шляхом локального застосування ШПФ до невеликих часових сегментів сигналу.

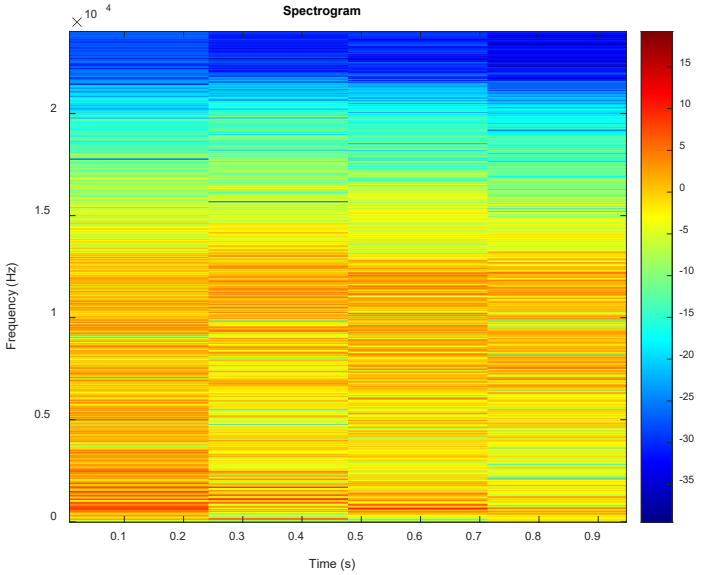


Рисунок 3.2 — Спектрограма МЛА *«DJI Mavic 3»*

Спектрограми забезпечують детальну візуалізацію частотного складу аудіосигналу в часі й можуть фіксувати перехідні та змінні в часі характеристики сигналу, тому зазвичай їх використовують для вилучення таких характеристик, як висота, ритм і тембр музичних звуків.

Вони підходять для завдань, які вимагають часово-частотної інформації. Вагомими недоліками спектрограм є їхня висока розмірність та менш компактне представлення, а також вони потребують ретельного розгляду співвідношення час-частота.

Мел-спектрограми — це спектрограми, де частоти (f) перетворені в шкалу мела (m) за такою формулою:

$$m = 2595 \cdot lg\left(1 + \frac{f}{700}\right) \tag{3.1}$$

Вони використовуються, як представлення ознак в аналізі аудіо, розпізнаванні мови, обробці музики та інших подібних завданнях. Кожен її стовпчик представляє величину частотного вмісту аудіосигналу в різні проміжки часу з акцентом на релевантних для сприйняття частотних мел-діапазонах (рис. 3.3).

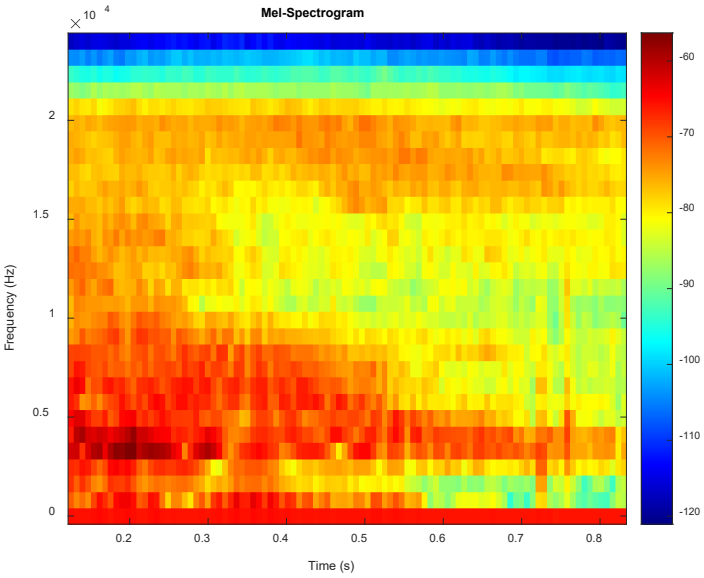


Рисунок 3.3 — Мел-спектрограма МЛА *«DJI Mavic 3»*

Вона зберігає більше спектральної інформації, ніж MFCC та краще підходить для завдань, які вимагають дрібнозернистої частотної інформації. Але розмірність мел-спектрограми в порівнянні з MFCC значно вища, а також є проблема обмеження часової розділової здатності через розмір кадру та вікна.

Інформацію про швидкість зміни спектральних смуг сигналу дає його кепструм — дискретне косинусне перетворення (ДКП) логарифма спектра часового сигналу. Отриманий спектр не лежить ні в частотній, ні в часовій області [37].

Мел-частотні кепстральні коефіцієнти (MFCC) — це коефіцієнти, що складають мел-частотний кепстр (рис. 3.4).

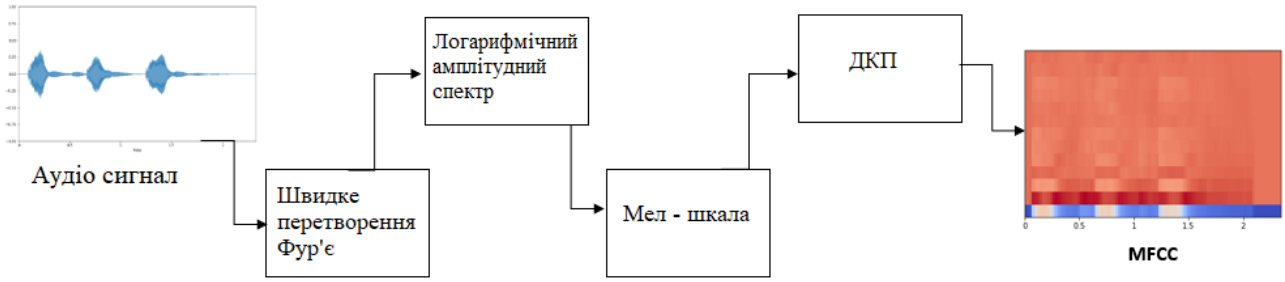


Рисунок 3.4 — Кроки для виділення MFCC з аудіосигналу

Кепструм передає різні значення, які будують форманти (характерний компонент якості мовного звуку) і тембр звуку тому вони, широко використовується в задачах розпізнавання мови, ідентифікації джерел звуку, класифікації аудіо та інших завдань, які передбачають аналіз спектрального вмісту сигналу [38]. MFCC стійкі до шуму та варіацій джерел, фіксують важливу спектральну інформацію та характеристики аудіосигналу, водночас зменшуючи розмірність простору ознак. З недоліків можна виділити втрати дрібнозернистих спектральних деталей за середніх обчислювальних ресурсів, тому вони менше підходять для завдань, де важлива точна інформація про частоту.

Щоб визначити MFCC потрібно до логарифмічно масштабованих виходів банку мел-фільтрів застосувати ДКП:

$$X(k) = \sum_{k=1}^{N-1} x(n) \cos\left(\frac{\pi n}{C}\left(k - \frac{1}{2}\right)\right), \quad k = 0, 1 \dots N-1 \qquad (3.2)$$

де *X(k)* — вихідна послідовність коефіцієнтів ДКП;

x(n) — вхідний аудіосигнал завдовжки N.

Вибір кількості мел-частотних кепстральних коефіцієнтів, що відповідають найбільш інформативним спектральним характеристикам аудіосигналу МЛА, має враховувати такі чинники, як характеристики сигналу, обчислювальні ресурси та компроміс між інформативністю і розмірністю. Оптимальна кількість MFCC може змінюватися залежно від конкретного завдання і визначається за допомогою експериментів та перевірок на валідаційному наборі даних. Орієнтовна загальна кількість кепстральних коефіцієнтів (*n_MFCC*) розраховується за наступною формулою:

$$n_MFCC = nBands + 2 \cdot k_{\delta} \tag{3.3}$$

де *nBands* — кількість банків мел-фільтрів (3.4),

k_{δ} — кількість додаткових дельта- або дельта-дельта-коефіцієнтів, що представляють собою відповідно першу та другу похідні від виразу 3.2.

Для отримання основних представлень аудіосигналу виконаємо техніку *зважування коефіцієнтів ШПФ за допомогою банків мел-фільтрів* — набору трикутних фільтрів, рівномірно розташованих уздовж шкали мел (рис 3.5).

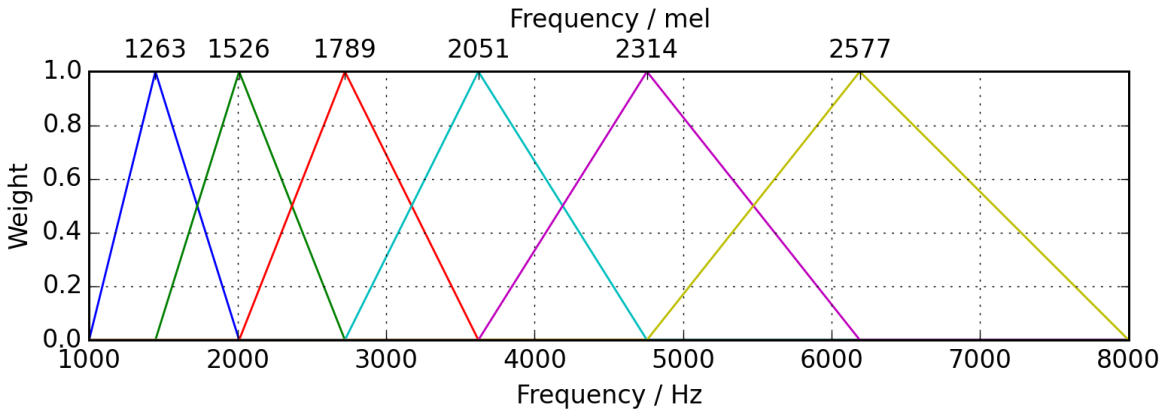


Рисунок 3.5 — Банк мел-фільтрів на шкалі мел

Їхню кількість треба вибирати таким способом, щоб збалансувати частотну роздільну здатність і обчислювальну складність аналізу. Занадто мала кількість фільтрів може призвести до поганої частотної роздільної здатності, тоді, як занадто велика їхня кількість, може збільшити обчислювальну складність без покращення частотної роздільної здатності. Тому зазвичай кількість банків мел-фільтрів обирають на основі наступних кроків:

1. Визначимо частотний діапазон, який потрібен для аналізу. Він напряму залежить від характеристик аудіосигналу та дорівнює інтервалу від 0 до частоти дискретизації (f_s) поділеній навпіл (за теоремою Найквіста).

2. Далі визначимо бажане значення частоти розділової здатності (f_{res}).

3. Вираз для обчислення кількісті банків мел-фільтрів *(numBands)* випливає з нелінійного перетворення 3.1 та бажаної частотної роздільної здатності:

$$nBands = 2595 \cdot \frac{lg\left(1+\frac{f_s}{2\cdot 700}\right)}{2\cdot f_{res}} - 1 \qquad (3.4)$$

Загальне емпіричне правило полягає у використанні кількісті фільтрів від 20 до 40, причому 26 або 30 є найпоширенішими варіантами.

Розрахуємо кількість банків фільтрів мел, що використовуються для аналізу аудіосигналів МЛА в розроблювальному алгоритмі. Діапазон частот аналізу знаходиться між 20 Гц та 20 кГц, що становить частотні межі роботи обраної моделі мікрофону. Бажана частотна розділова здатність 50 Гц, тому за формулою 3.4 кількість мел смуг дорівнює 30.

Для зважування коефіцієнтів ШПФ кожен трикутний фільтр у банку фільтрів мел потрібно згорнути зі спектром, отриманим у результаті ШПФ. Вихідні дані банку фільтрів обчислюються, як зважена сума значень коефіцієнтів ШПФ у межах частотного діапазону кожного фільтра.

Щоб знайти частотний діапазон кожного мел-фільтру ($range_{Mel}$), необхідно розділити шкалу мел на рівні інтервали (N):

$$range_{Mel} = \frac{m}{nBands + 1} \cdot N, \quad N = 1 \dots nBands + 1 \qquad (3.5)$$

А потім перевести мел-діапазон назад у лінійну шкалу частот, використовуючи зворотну формулу для шкали мел:

$$f = 700 \cdot \left(10^{\frac{range_{Mel}}{2595}} - 1\right) \qquad (3.6)$$

Знаючи частотні межі роботи обраної моделі мікрофону з 3.1 випливає, що весь мел-частний діапазон m знаходиться в проміжку від *1* до *3817* Мел.

У якості прикладу, розрахуємо межі частот першого діапазону ($N=1$) в алгоритмі аналізу сигналів. Щоб визначити нижню межу, у 3.5 підставимо

значення виразу 3.1 (*1* Мел) за мінімальної робочої частоти мікрофону (*20* Гц) та потрібну кількість банків мел-фільтрів, обрану за 3.4 (*nBands = 30*). Отже, нижня межа дорівнює *1* Мел або *0.64* Гц. Аналогічну послідовність дій проведемо і для розрахунку верхньої межі, але вже за максимальної частоти роботи мікрофону (*20* кГц). Відповідно $range_{Mel}$ ($N = 1$) дорівнює *123* Мел, приблизно 81 Гц (формула 3.6). У такий спосіб можна легко знайти частотний діапазон кожного мел-фільтру, підставляючи в 3.5 бажаний номер фільтра (*N*).

Виходи банку фільтрів часто логарифмічно масштабуються за допомогою логарифма або функції логарифмічного стиснення. Таке масштабування допомагає наблизитися до нелінійного частотного сприйняття гучності й забезпечує стиснення динамічного діапазону спектра енергії.

Отже, найбільш ефективними методами фіксації локальних характеристик, що ляжуть у основу роботи алгоритму класифікації та розпізнавання сигналів від дронів, є мел-спектрограми та MFCC.

3.4. Підготовка наборів навчальних даних

Після етапів попередньої обробки аудіосигналів МЛА та шумів оточуючого середовища, перетворення їх у формат, придатний для подачі на вхід НМ, вихідний набір даних необхідно розділити на три набори: дані *навчання*, *перевірки* (валідації) та *тестування*. Співвідношення розподілу залежить від розміру набору даних, тому стандартною практикою зазвичай є розділення 60/20/20 або 70/15/15 [36].

Навчальний комплект застосовується для навчання нейронної мережі, яка буде класифікувати дані, за допомогою завантаження розмічених прикладів аудіо шумів різних моделей МЛА.

Набір *перевірки (валідації)* — це набір даних, що використовуються для налаштування гіперпараметрів моделі, топології НМ та запобігання переобладнанню. Під час процесу навчання продуктивність моделі оцінюється на перевірочному наборі після кожної епохи, а гіперпараметри відповідно коригуються. Дані перевірки грають вагому роль у прийнятті рішення про

припинення навчання контрольованої НМ, вони допомагають збалансувати компроміс між переобладнанням (підгонкою моделі надто близько до навчальних даних) і недостатнім припасуванням (нездатність зафіксувати закономірності в даних).

Дані *тестування* — це набір даних, отриманих із процесу навчання та використаних для оцінювання моделі навченого класифікатора. Вони використовується для вимірювання точності моделі та її здатності узагальнювати нові, невідомі дані. Дані тестування використовуються один раз, наприкінці процесу навчання, для отримання остаточної оцінки продуктивності моделі.

3.5. Створення та навчання моделі нейронної мережі

Застосування класичних НМ для реалізації завдання розпізнавання та класифікації аудіосигналів ускладняється, як правило, великою розмірністю вектору вхідних значень мережі, значною кількістю нейронів у прихованих шарах, що призводить до серйозних витрат обчислювальної потужності на навчання НМ. Але згортковим нейронним мережам (ЗНМ) описані вище недоліки властиві в меншому ступені [39], тож зупинимо свій вибір на них.

Згорткові нейронні мережі — клас НМ прямого поширення, що використовують алгоритми глибокого навчання (*Deep Learning*) для ефективного аналізу різних типів сигналу з будь-якого давача та вимагають мінімальний обсяг попередньої обробки вхідних даних через спеціальну архітектуру використання багатошарових перцептронів — кібернетичної моделі для імітації машинами процесу сприйняття інформації людським мозком.

Будь-яка штучна НМ у своєму найпростішому представленні складається з впорядкованого набору таких компонентів: вхідного та вихідного шару нейронів, а також проміжних прихованих або обчислювальних шарів (рис. 3.6).

Ключовим моментом під час проектування обчислювальних шарів ЗНМ є поняття *«спільної» ваги*, коли частина нейронів деякого розглянутого шару НМ використовує однакові вагові коефіцієнти. Такі нейрони групуються в мапу ознак, де кожен нейрон з’єднаний із групою нейронів попереднього шару.

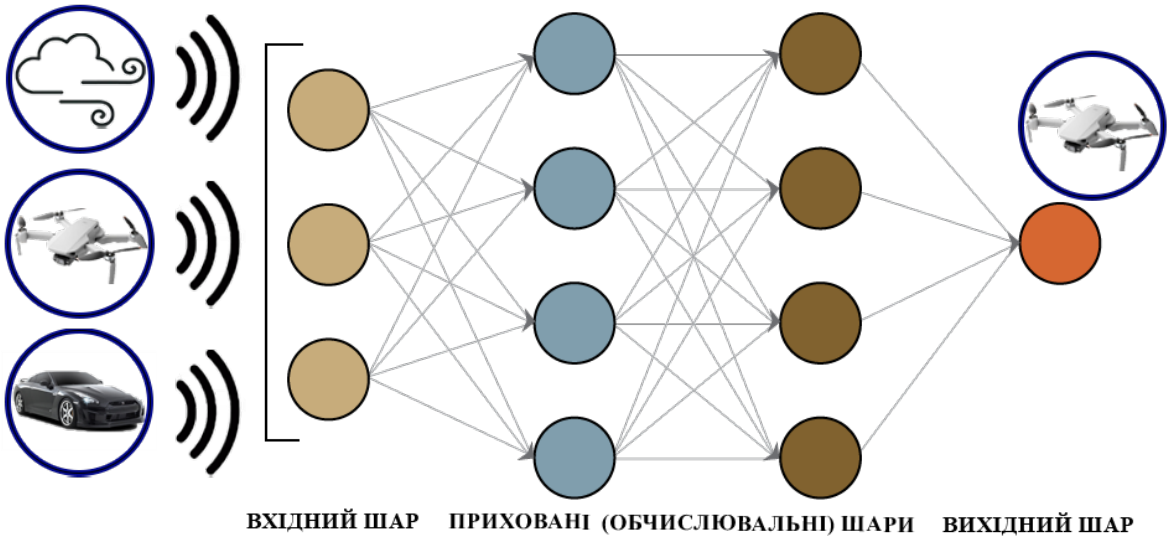


Рисунок 3.6 — Спрощена топологія НМ для виявлення МЛА

Під час обчислення мережі кожен із нейронів попередніх шарів виконує операцію згортки з певною ділянкою — фільтром або Кернелом, який представляє собою невелику сітку заданої ширини і висоти, та завдовжки, що простягається через усю глибину вхідного уявлення (рис. 3.7). Побудовані за таким принципом шари НМ назвали *згортковими*. Завдяки такому підходу зменшується об'єм необхідної пам'яті внаслідок мінімізації кількості вільних параметрів вхідних нейронів та поліпшується продуктивність алгоритму.

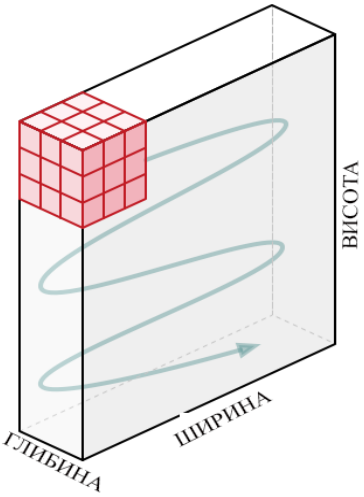


Рисунок 3.7 — Згортковий шар 3D НМ

Варто зауважити, що крайні пікселі входять у меншу кількість згорток, ніж ті, що лежать всередині. Саме тому, у згортковому шарі часто виконується доповнення країв зображення «фейковими» пікселями — функція ***падингу***. Виходи з попередніх шарів доповнюються пікселями (зазвичай рівними 0 — «*нульовий падинг*») так, щоби після згортки зберігався розмір даних. Отже, під час проходження по елементах, ядро дає змогу справжнім крайнім пікселям опинитися у своєму центрі, а далі поширюється на підроблені пікселі за межами оригінального краю, створюючи вихідну матрицю того ж розміру, що і вхідна.

Такі згортки називають однаковими (*same convolution*), а якщо ж доповнення не відбувається, тоді згортку називають правильною (*valid convolution*).

Поміж способів доповнення даних крайніми пікселями, виділяють такі:

- Доповнення нулями (*zero shift*): 00[RTF]00;
- Подовження границі крайніми елементами (*border extension*): RR[RTF]FF;
- Дзеркальне доповнення (*mirror shift*): TR[RTF]FT;
- Доповнення по колу (*cyclic shift*): TF[RTF]RT.

Ще одним важливим параметром шару згортки є ***страйдинг*** (*stride*) або ***зсув***. Зазвичай операція згортки виконується послідовно для кожного з пікселів, однак іноді застосовується зсув, відмінний від одиниці — скалярний добуток розраховується не з усіма допустимими положеннями ядра, а вже з кратними зсуву s (рис. 3.8). Тоді, якщо вхід мав розмірність $w \times h$, а ядро згортки — $k_x \times k_y$ та застосовувався зсув s, то вихід буде мати розмірність:

$$\left[\frac{w - k_x}{s} + 1\right] \times \left[\frac{h - k_y}{s} + 1\right]$$

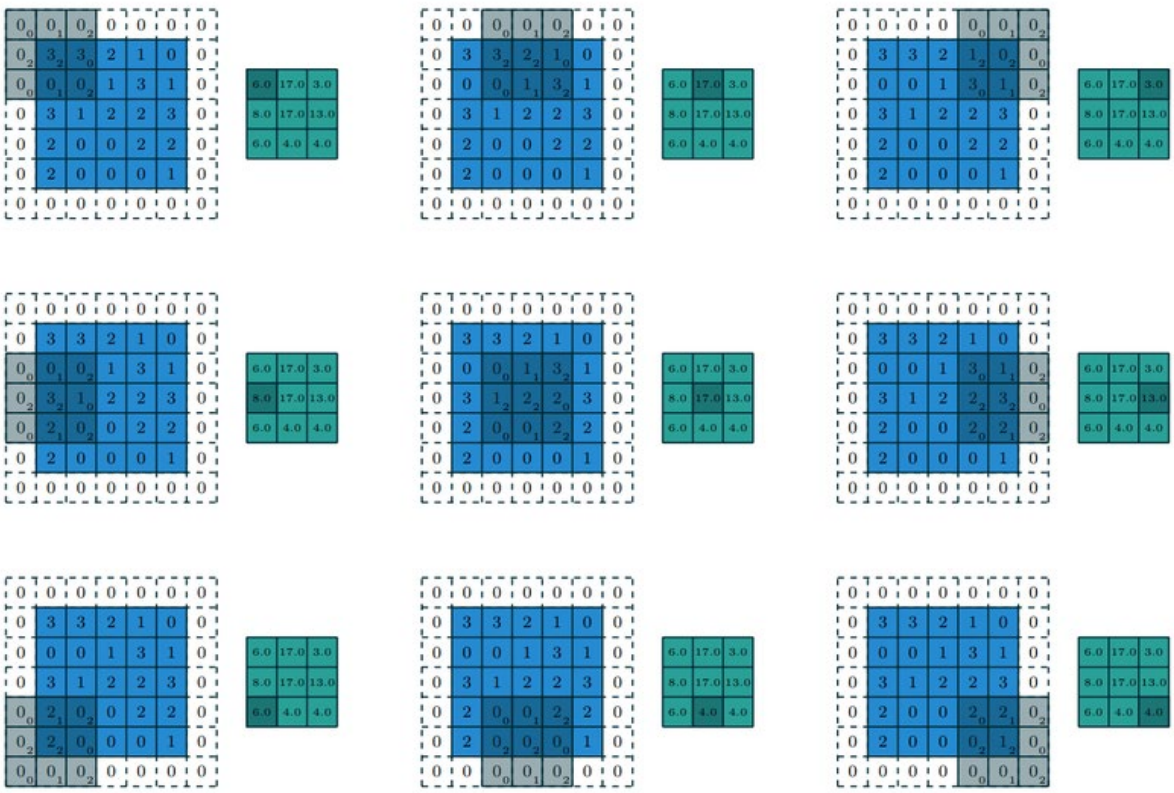


Рисунок 3.8 — Функція падингу зі страйдингом 2 під час операції згортки

Функція страйдингу застосовується для зменшення просторової розмірності за великої кількості каналів, простіше кажучи, щоб здобути інформацію на виході значно меншого розміру, ніж на вході.

Також у прихованих шарах ЗНМ можуть бути шари *субдискретизації* або ***пулінгу***, де виконується операція зменшення просторового об'єму мапи ознак, що дає змогу значно знизити обчислювальну потужність, необхідну для оброблення даних. Крім того, це важливо і для отримання домінуючих ознак, інваріантного обертання та позиціонування, відповідно підтримуючи хід ефективного навчання моделі.

Існує два основних типа пулінгу: максимальний (*Max Pooling*) та середній (*Average Pooling*). Максимальний пулінг повертає найбільше значення ваги фрагменту зображення, що покривається Кернелом, а середній — середнє з усіх значень частини зображення (рис.3.9).

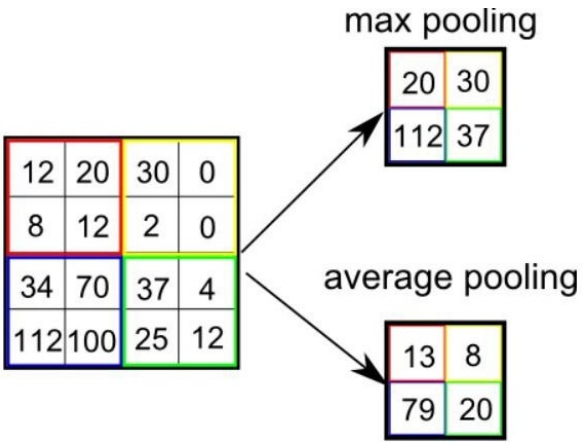


Рисунок 3.9 — Основні типи пулінгу

Максимальний пулінг також виступає в ролі шумоподавлювача (*Noise Suppressant*). Він повністю виключає шумові сигнали, також поєднує придушення шуму зі зменшенням розмірності. Середній пулінг просто використовує зменшення розмірності, як спосіб придушення шуму. Тобто, можна сказати, що Max Pooling працює набагато краще, ніж Average Pooling.

Згортковий та пулінговий шари разом утворюють і-й шар ЗНМ. Залежно від складності вхідних даних кількість таких шарів, може бути збільшено для того, щоби більш точно захопити деталі низького рівня, але також збільшується і обчислювальна потужність. Виконання описаного вище процесу дає змогу моделі успішно вивчити ознаки.

У *повнозв'язкових* шарах проходить процес з'єднання нейронів одного шару з нейронами наступного для згладжування та представлення вихідних даних, що зведені у вектор-стовбець. Згладжений висновок подається в НМ із

прямим зв'язком, і потім до кожної ітерації навчання застосовується зворотне поширення. Після серії епох модель набуває здібність відрізняти домінуючі та низькорівневі ознаки, а також класифікувати їх використовуючи функції *Softmax*.

Основною *ідеєю* ЗНМ є можливість чергування в довільному порядку всіх видів прихованих шарів, що дає змогу будувати мапу ознак із мап ознак. На практиці це дає можливість ЗНМ розрізнювати важкі ієрархії ознак [40].

Зазвичай **навчання** НМ проходить у 2 стадії:

- *Пряме поширення помилки* (рис. 3.10)

Для побудови шару обчислення вхідні дані множаться на початкові ваги (w_1,w_2), що задають рандомно:

$$h_1 = (x_1 \cdot w_1) + (x_2 \cdot w_1)$$

$$h_2 = (x_1 \cdot w_2) + (x_2 \cdot w_2)$$

$$h_3 = (x_1 \cdot w_3) + (x_2 \cdot w_3)$$

Результат на виході мережі (y_1) отримують через пропускання даних з обчислювального шару крізь нелінійну функцію активації: $y_1 = fn(h_1, h_2, h_3)$

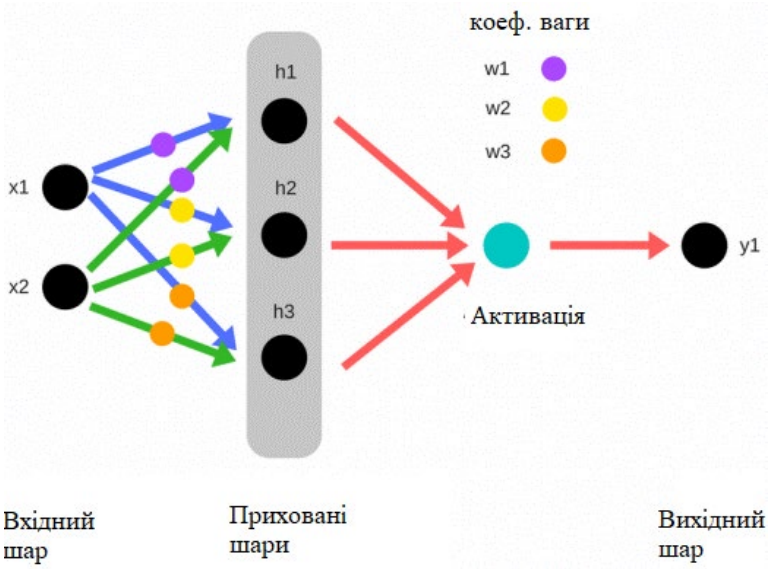


Рисунок 3.10 — Приклад прямого розповсюдження помилки в ЗНМ

- *Зворотне поширення помилки* (рис. 3.11).

Загальна помилка визначається, як різниця між вихідним значенням «*y*», прогнозованим із навчального набору та «y_1» отриманим на фазі прямого поширення помилки. Усі значення так само пропускаються крізь функцію втрат.

Часткова похідна помилки розраховується за кожною вагою (ці часткові диференціали показують внесок кожного коефіцієнту ваги в загальну помилку (*total_loss*)). Далі диференціали множаться на коефіцієнт швидкості навчання η і віднімаються з відповідної ваги.

Оновлені значення коефіцієнтів ваги такі:

$$w_1 = w_1 - \left(\eta \cdot \frac{\partial(err)}{\partial(w_1)}\right)$$

$$w_2 = w_2 - \left(\eta \cdot \frac{\partial(err)}{\partial(w_2)}\right)$$

$$w_3 = w_3 - \left(\eta \cdot \frac{\partial(err)}{\partial(w_3)}\right)$$

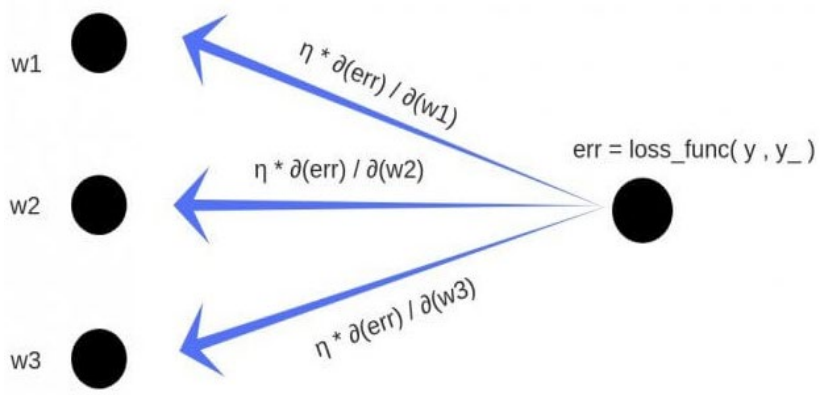


Рисунок 3.11 — Зворотне поширення помилки

Під час прямого поширення відбувається передбачення відповіді, а в разі зворотного — зменшується помилка між прогнозованою та фактичною відповіддю.

Функція активації (передавальна функція) — важливий інструмент, який визначає функціональні можливості НМ: які нейрони будуть активовані, яка інформація буде передаватися наступним шарам та метод навчання мережі.

Щоб глибокі НМ не втрачали здібності до якісного навчання, функція активації мусить мати такі властивості:

- *Нелінійність* — підвищує ступінь свободи, що зволяє узагальнити проблему високої розмірності, завдяки чому двошарова НМ стає універсальним апроксиматором функції [41]. Тотожна ПФ не задовольняє такій властивості. Якщо декілька рівнів прихованих шарів застосовують тотожну функцію, тоді НМ рівнозначна одношаровій моделі.

- *Неперервну диференційованість* — властивість для забезпечення методів оптимізації на базі градієнтного спуску (крім функції ReLU).

- *Монотонність* — якщо функція активації монотонна, поверхня помилок, що асоціюється з однорівневою моделлю, гарантовано буде опуклою [42].

- *Зона визначення* — якщо купа значень функції активації обмежені, методи навчання на основі градієнту більш стабільні, оскільки представлення еталонів істотно впливають, лише на обмежений набір ваги зв’язків. Якщо зона значень нескінченна, навчання, зазвичай, більш ефективне, оскільки представлення

еталонів впливає на більшість вагових коефіцієнтів. В останньому випадку зазвичай необхідна менша швидкість навчання.

- *Гладкість функції з монотонною похідною* — забезпечує більш високий ступінь спільності.

- *Апроксимація* біля точки [0;0] *до тотожної функції* $f(x) = x$. Якщо функції активації притаманна ця властивість, НМ буде навчатись ефективніше, якщо її ваги ініціалізовані малими рандомними значеннями. Якщо ж функція активації не апроксимує тотожність біля початку координат, необхідно бути обережним під час ініціалізації ваг [43].

У таблиці 3.1 наведено перелік найпоширеніших функцій активації для ЗНМ, проаналізуємо їхні основні переваги та недоліки.

Функція **сигмоїда** перетворює значення, що надходять у неї, у дійсний діапазон [0, 1]. Тобто, якщо вхідні дані матимуть великі додатні значення, то після перетворення вони дорівнюватимуть приблизно одиниці, а від'ємні числа будуть наближатися до нуля. Це доволі популярна функція, яку можна інтерпретувати, як частоту збудження нейрона.

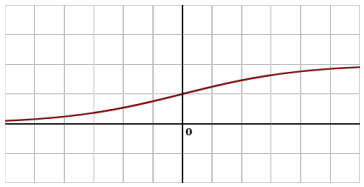
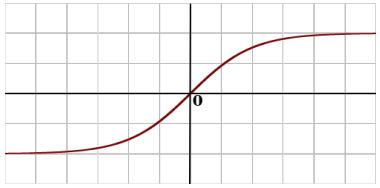
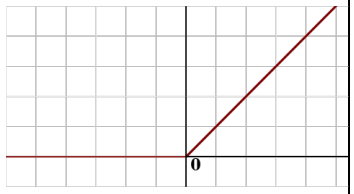
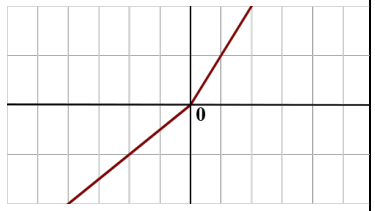
Але функція сигмоїда має певні недоліки:

1. *Насичені нейрони можуть «придушити» градієнт*. Візьмемо сигмоїдальний вузол обчислювального графа та завантажимо в нього вхідні дані X. Коли відбувається зворотній прохід, градієнт, що сходить, дорівнює $dL/d\sigma$, а локальний — $dL/d\sigma * d\sigma/dx$.

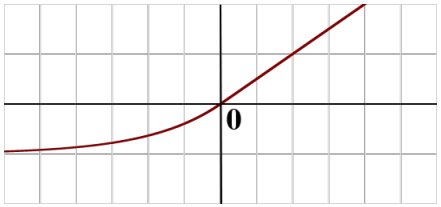
Наприклад, коли $X = -10$, то градієнт стане нульовим, оскільки великі від'ємні значення лежать на прямій ділянці функції сигмоїда. Отже, у всі наступні вузли будуть передаватися нульові похідні — це і «придушує» градієнтний потік.

У випадку, коли $X = 0$ все гаразд, як і для інших близьких до нуля значень. А ось за умови $X = 10$ градієнт знову обнулиться. Тому що сигмоїда не працює для надто великих додатних чи від'ємних даних, що і видно з характеру поведінки її графіку.

Таблиця 3.1 — «Передавальні функції для ЗНМ»

Назва ф-ії	*Сігмоїд (Sigmoid)*	*Гіперболічний тангенс (tanh)*	*Випрямлена лінійна (ReLU)* [44]	*Leaky ReLU* [45]
Графік				
Рівняння	$f(x) = \sigma(x) = \frac{1}{1+e^{-x}}$	$f(x) = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$	$f(x) = \begin{cases} 0 & \text{при } x < 0 \\ x & \text{при } x \geq 0 \end{cases}$	$f(x) = \begin{cases} 0.01x & \text{при } x < 0 \\ x & \text{при } x \geq 0 \end{cases}$
Похідна (за x)	$f'^{(x)} = f(x)(1 - f(x))$	$f'^{(x)} = 1 - f(x)^2$	$f'(x) = \begin{cases} 0 & \text{при } x < 0 \\ 1 & \text{при } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0.01 & \text{при } x < 0 \\ 1 & \text{при } x \geq 0 \end{cases}$
Зона	$(0,1)$	$(-1,1)$	$[0,\infty)$	$(-\infty,\infty)$
Порядок гладкості	C^∞	C^∞	C^0	C^0
Монотонність	+	+	+	+
Монотонність похідної	–	–	+	+
Апроксимація до тотожної ф-ії біля т. [0;0]	–	+	–	–

Продовження таблиці 3.1.

Назва ф-ії	*ELU* [46]	*Maxout* [47]	*Softmax*
Графік			
Рівняння	$f(\alpha, x) = \begin{cases} \alpha(e^x - 1) & \text{при } x < 0 \\ x & \text{при } x \geq 0 \end{cases}$	$f(\vec{x}) = \max\,(w_1^T x + b_1, w_2^T x + b_2)$	$f_i(\vec{x}) = \frac{e^{x_i}}{\sum_{j=1}^{J} e^{x_j}}$ для $i = 1, \ldots J$
Похідна (за x)	$f'(\alpha, x) = \begin{cases} f(\alpha, x) + \alpha & \text{при } x < 0 \\ 1 & \text{при } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{при } x < 0 \\ 1 & \text{при } x \geq 0 \end{cases}$	$\frac{\partial f_i(\vec{x})}{\partial x_j} = f_i(\vec{x})(\delta_{ij} - f_j(\vec{x}))$ [42]
Зона	$(-\alpha, \infty)$	$(-\infty, \infty)$	(0,1)
Порядок гладкості	$\begin{cases} C^1 & \text{коли } \alpha = 1 \\ C^0 & \text{інакше} \end{cases}$	C^0	C^∞
Монотонність	$+$ $(\alpha \geq 0)$		
Монотонність похідної	$+$ $(0 \leq \alpha \leq 1)$		
Апроксимація до тотожної ф-ії біля т. [0;0]	$-$ $(\alpha = 1)$		

2. *Вихідні значення сигмоїда не центровані нулем*. Нехай початкові дані будуть повністю додатні, тоді під час зворотного поширення всі градієнти будуть або додатними, або від'ємними (залежно від градієнта *f*). Це призведе до того, що всі ваги в разі оновлення також будуть або збільшені, або зменшені, а градієнтний потік стане зигзагоподібним. Тому із самого початку варто впоряджати дані так, щоб їхнім середнім значенням був нуль.

3. *Функцію експоненти досить дорого рахувати*. Це не дуже істотна проблема, оскільки скалярні добутки під час згортки витрачають набагато більше обчислювальної потужності, але під час порівняльного аналізу з іншими передавальними функціями це варто зауважити.

Тангенс, на відміну від сигмоїди, має дві основні відмінності: перетворює дані в діапазон [-1, 1] та має нульове центрування, що виключає другу проблему сигмоїди. Значення градієнта під час зворотного поширенні, як і раніше, можуть обнулятись, проте краще використовувати тангенс.

ReLU *(Rectified Linear Unit)* обчислює функцію $f(x) = \max(0, x)$, тобто просто видає значення «нуль» або «не нуль». Це розв'язує проблему обнуління градієнта для додатних чисел. Крім того, ReLU простіша та приблизно в 6 разів швидша в обчислюванні за сигмоїди й тангенси. Однак у ній також немає нульового центрування. Іншим вагомим недоліком є «придушення» градієнту за від'ємних вхідних даних. Це може призвести до того, що половина нейронів буде неактивною та не зможе оновлюватись.

Проблему можна вирішити, задавши нижчу швидкість навчання та підібравши інші вагові коефіцієнти. Або використовувати модифікації ReLU.

Відмінність функції **Leaky ReLU** в тому, що вона має невеликий нахил у лівій напівплощині — за від'ємних вхідних даних градієнт буде не нульовий. Водночас її легко обчислити, відповідно вона вирішує практично всі перелічені проблеми. Одним із її різновидів є PReLU, яка виглядає як $f(x) = \max(\alpha x, x)$.

Функція **ELU** схожа на Leaky ReLU та має всі її переваги, але включає в себе експоненту, що робить її обчислення дорожчим. Її варто використовувати в тих випадках, коли вам важлива стійкість до шумових даних.

Maxout вибирає максимальну суму з двох наборів ваг, помножених на початкові дані з урахуванням зміщення. Тим самим він узагальнює ReLU та Leaky ReLU, не обнуляючи градієнт. Але maxout вимагає подвоєння параметрів і нейронів.

Softmax або нормована експоненційна функція використовується для завдання класифікації в глибоких НМ за кількості класів більше, ніж два. Softmax перетворює вектор $\vec{x}$, що має розмірність J, у вектор $\vec{f}$ такого ж розміру, де кожна з координат f_i отриманого вектору представлена дійсним числом у проміжку [0,1], а сума координат дорівнює 1. Координати f_i отриманого вектору водночас трактуються, як ймовірності того, що об'єкт належить до класу i. Для навчання таких НМ, як функцію втрат зазвичай використовують перехресну ентропію.

Підсумовуючи вище сказане, в алгоритмі глибокого навчання НМ краще використовувати функцію активації ReLU, завдяки своїй простоті, обчислювальній ефективності та здатності вносити нелінійність у мережу, що дає змогу створювати більш потужні та виразні моделі.

3.5.1. Архітектура моделі згорткової нейронної мережі

Як було описано вище, проста згорткова нейронна мережа — це набір шарів, кожен із яких перетворює одне подання на інше з використанням певної диференційованої функції. Використаємо три головні типи шарів для побудови ЗНМ: згортковий шар, шар підвибірки та повнозв'язний шар (такий самий, який використовується у звичайній НМ). Щоб отримати архітектуру згорткової нейронної мережі для класифікації аудіосигналів малих безпілотних літальних апаратів, розташуємо всі шари послідовно (рис. 3.12). Побудуємо отриману модель НМ використовуючи програмне середовище MATLAB.

Вхідний шар (*Input Layer*) НМ приймає всі вхідні сигнали та передає далі на наступні рівні. Варто зазначити, що зазвичай цей шар очікує вхідні дані у вигляді зображення, тому для аналізу аудіосигналів їх попередньо необхідно представити в графічному вигяді, як двовимірне зображення — часо-частотне представлення за допомогою спектрограми або Мел-спектрограми.

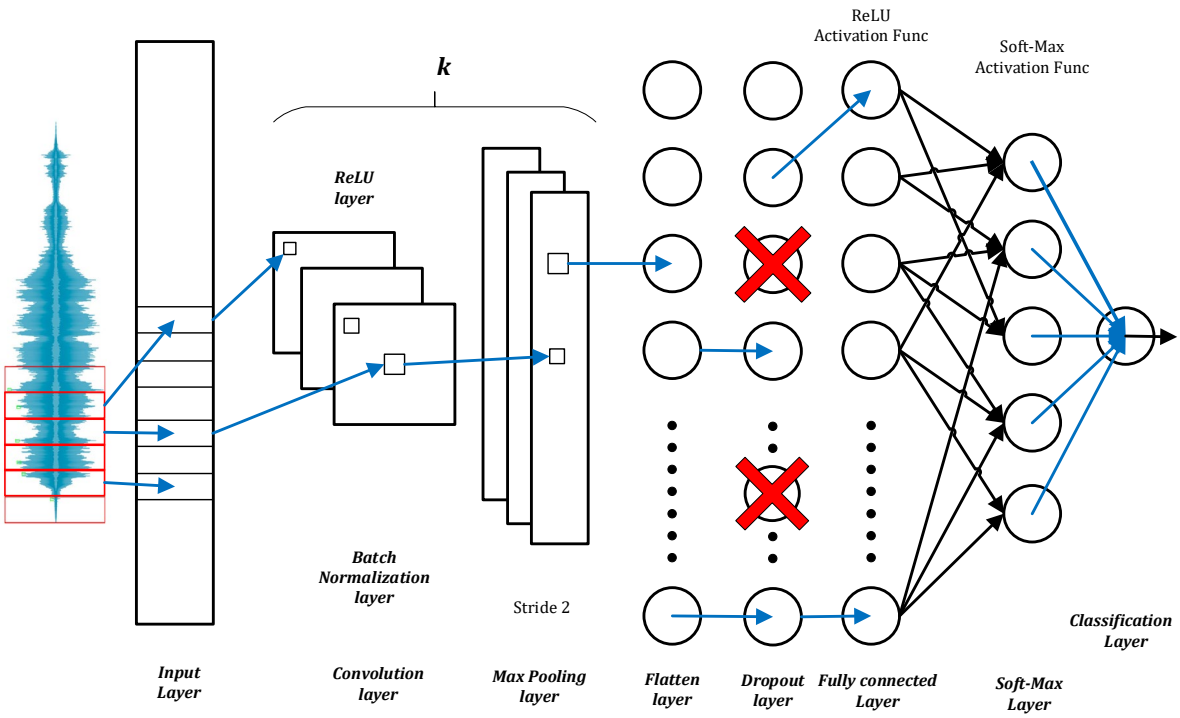


Рисунок 3.12 — Архітектура ЗНМ для класифікації аудіосигналів МЛА

У програмному середовищі MATLAB є універсальний вхідний шар *imageInputLayer*, який можна використовувати для аналізу аудіофайлів. Він приймає дані у вигляді зображення, де кожен піксель представляє певний атрибут аудіосигналу, що забезпечує можливість зручного використання існуючих інструментів і методів аналізу зображень для аудіоданих.

```
imageInputLayer(inputSize)
```

Розмір вхідних даних (*inputSize*) задається у вигляді вектора-рядка із цілих чисел *[numHops nBands numChannels]*, де *numHops* — кількість часових інтервалів в аудіофайлі, *nBands* — кількість банків фільтрів мел для представлення аудіосигналу, а *numChannels* — кількість аудіоканалів (дорівнює 1, оскільки використовується лінійний вхід звукової карти).

Далі необхідно створити **<u>шари згортки</u>** (*Convolution Layer*) та задати їм відповідні параметри, для цього скористаємось такою функцією:

```
convolution2dLayer(filterSize,numFilters,Name=Value)
```

Першим параметром необхідно задати розмір фільтра (***filterSize***), який визначає розмір сприйнятливого поля шару згортки. Найпоширеніші розміри фільтрів: 3x3, 5x5 або 7x7. Менші розміри фільтрів захоплюють дрібні деталі, а

більші розміри — глобальні особливості. Вибір розміру фільтра залежить від складності даних і бажаного рівня абстракції. Зазвичай для ефективного аналізу складових аудіосигналу використовують фільтр 3x3.

Другим важливим параметром є кількість фільтрів (***numFilters***), яка визначає глибину або кількість каналів у вихідній карті ознак. Вибір кількості фільтрів у шарі згортки передбачає пошук відповідного балансу між складністю моделі та її продуктивністю. Щоб обрати кількість фільтрів у шарах згортки, потрібно зважати на наступні фактори:

1) Вхідні та вихідні розміри

Кількість фільтрів у згортковому шарі відповідає кількості вихідних каналів або карт ознак, створених цим шаром. Зазвичай це залежить від розміру вхідних даних і бажаного розміру вихідних карт ознак. Кількість фільтрів у першому шарі часто вибирають у діапазоні від 10 до 30.

2) Пропускна здатність і складність моделі

Збільшення кількості фільтрів у шарі згортки збільшує здатність моделі вивчати складні об'єкти, дає змогу мережі вловлювати дрібніші деталі та закономірності у вхідних даних. Однак більша кількість фільтрів також збільшує кількість параметрів у мережі, що може призвести до довшого часу навчання та більшого ризику перенавчання, особливо, якщо набір даних невеликий.

3) Глибина мережі

Кількість фільтрів збільшується в міру заглиблення в мережу. Це відомо як «пірамідальна» архітектура, яка надає можливість мережі поступово вивчати представлення більш високого рівня, збільшуючи кількість фільтрів. Глибші шари можуть вивчати більш абстрактні ознаки, побудовані на низькорівневих ознаках, вивчених у попередніх шарах. Простою стратегією для збільшення потужності моделі є подвоєння кількості фільтрів у кожному наступному шарі.

Визначимо кількість фільтрів у згорткових шарах розроблюваної моделі. Спроектуємо ЗНМ за «пірамідальною» архітектурою, у першому шарі згортки візьмемо середнє значення діапазону початкової кількості каналів фільтрів

(*numFilters = 15*), а в кожному наступному шарі, поступово збільшуємо їхню кількість у *2n* разів, де *n* — номер згорткового шару.

Точність моделі та її здатність до узагальнення оцінюється на валідаційному наборі даних. Якщо модель надмірно пристосована, доведеться зменшити кількість фільтрів. Якщо ж модель недостатньо пристосована або намагається вивчити складні ознаки, потрібно розглянути можливість збільшення кількості фільтрів.

Останнім необов’язковим параметром функції зазначимо падинг (*«Padding»*), що може мати значення «*same*» (однаковий) або «valid» (дійсний). Параметр «Padding», встановлений на «same», означає, що вхідні дані доповнюються нулями так, щоб вихідні дані мали такі ж просторові розміри, що і вхідні. Значення «valid» не додає жодного доповнення, і вихідні карти ознак мають зменшені просторові розміри. Вибір типу заповнення залежить від бажаного розміру вихідної карти і від того, чи важливим є збереження просторової інформації.

Отже, кожен згортковий шар мережі буде складатись із *2n·numFilters* фільтрів розміром *3x3* та падингу зі значенням «*same*», щоб вихідні та вхідні карти об'єктів мали аналогічні просторові розміри. Цей шар допомагає вловити локальні закономірності та особливості вхідних даних.

Щоб дати нагоду мережі вивчати більш складні представлення, введемо нелінійність за допомогою шару з функцією активації випрямленої лінійної одиниці **ReLU** (*ReLU Layer*). Вона є обчислювально ефективною, оскільки просто встановлює від'ємні значення виходу попереднього шару в нуль без будь-яких складних обчислень, а додатні залишає незмінними. Це перетворення не змінює розмірність даних. ReLU сприяє розрідженим представленням, що може допомогти зробити мережу більш стійкою до шуму і зменшити перенавчання, зосередившись на найбільш релевантних ознаках.

Для прискорення навчання ЗНМ та зменшення чутливості до мережевої ініціалізації, між згортковими шарами та передавальною функцією ReLU, застосуємо **шар пакетної нормалізації** (*Batch Normalization Layer*).

Метою пакетної нормалізації є розв’язання проблеми внутрішнього коваріаційного зсуву — зміні розподілу входів шару, яка відбувається під час навчання, коли оновлюються параметри попередніх шарів. Цей зсув може зробити процес навчання мережі більш повільним та складним.

Шар пакетної нормалізації нормалізує активації шару за допомогою віднімання середнього значення пакету й ділення на стандартне відхилення пакету. Це гарантує, що активації будуть мати нульове середнє значення та одиничну дисперсію, ефективно зменшуючи внутрішній зсув коваріацій.

Перевагами використання пакетної нормалізації є:

1) Покращена швидкість навчання та збіжність

Зменшуючи внутрішній зсув коваріацій, пакетна нормалізація може пришвидшити процес навчання, що дасть змогу мережі швидше збігатися.

2) Кращий градієнтний потік

Нормалізація вхідних даних шару допомагає пом'якшити проблеми зникнення та вибуху значень градієнта, що полегшує навчання глибоких мереж.

3) Ефект регуляризації

Пакетна нормалізація діє, як форма регуляризації, додаючи невелику кількість шуму до входів шару під час навчання, що може допомогти запобігти надмірному пристосуванню.

4) *Підвищена стійкість до ініціалізації та вибору гіперпараметрів*

Пакетна нормалізація робить мережу більш стійкою до початкових значень параметрів і менш чутливою до швидкості навчання та інших налаштувань гіперпараметрів. Однак важливо зазначити, що вона вводить додаткові гіперпараметри, наприклад, імпульс для обчислення пакетної статистики, і вимагає ретельного налаштування для оптимальної продуктивності.

Шар максимального об'єднання (*Max Pooling Layer*) застосовується для поступового зменшення просторового розміру вхідних даних, зменшуючи кількість параметрів і обчислювальну складність, водночас зберігаючи найважливіші ознаки. Він працює з кожною картою ознак незалежно, витягує максимальне значення в межах певного вікна або розміру пулу, допомагає

виділити найбільш домінуючі ознаки й забезпечує певну інваріантність перекладу, роблячи мережу більш стійкою до невеликих варіацій вхідних даних.

Операція максимального об'єднання зазвичай застосовується з фіксованим кроком, що визначає величину, на яку зміщується вікно об'єднання після кожної операції об'єднання. Найпоширенішим підходом є використання розміру об'єднання *2x2* або *3x3* і кроку страйдингу *2*, що призводить до зменшення просторових розмірів карт об'єктів удвічі. Переваги використання шару максимального об'єднання полягають у такому:

1) Зменшення розмірності

Зменшуючи вибірку карт ознак, шар максимального об'єднання зменшує просторові розміри, що допомагає контролювати надмірне припасування і зменшує обчислювальну складність наступних шарів.

2) *Інваріантність трансляції*

Максимальне об'єднання забезпечує певний ступінь інваріантності трансляції, тобто, навіть якщо об'єкт трохи зміщується в межах вікна об'єднання, максимальне значення однаково фіксується, зберігаючи важливу інформацію.

3) *Стійкість до невеликих просторових змін*

Максимальне об'єднання допомагає мережі бути менш чутливою до невеликих зсувів або викривлень вхідних даних, роблячи модель більш стійкою до шуму та невеликих варіацій.

Однак важливо зазначити, що максимальне об'єднання може призвести до певної втрати просторової інформації, оскільки зберігається лише максимальне значення в межах кожного вікна об'єднання. У деяких випадках ця втрата інформації може бути компромісом між зменшенням обчислювальної складності та збереженням дрібних деталей. Для усунення цього недоліку можна дослідити альтернативні методи об'єднання, такі, як середнє або адаптивне об'єднання.

У програмному алгоритмі шар максимального об'єднання задається наступним чином:

```
maxPooling2dLayer(3,Stride=2,Padding="same")
```

Параметр *«3»* задає зону об'єднання *3x3*, а параметр *«Stride»,* встановлений на 2, означає, що операція об'єднання пропускає кожен другий елемент. Параметр «*Padding*» зі значенням «*same*» показує, що вихідні та вхідні дані будуть однакового розміру.

У ЗНМ згорткові шари та максимального об'єднання часто чергуються між собою для ефективного ієрархічного вивчення ознак та зменшення просторових розмірів. Згорткові шари витягують локальні шаблони та особливості з вхідних даних, а шари максимального об'єднання агрегують найпомітніші ознаки і зменшують просторові розміри, що дає змогу мережі виявляти схожі шаблони в різних місцях, захоплюючи ознаки вищого рівня. Чергуючи ці шари, ЗНМ можуть вивчати дедалі складніші й абстрактніші ознаки.

Визначення кількості пар шарів згортки та максимального об'єднання (***k***) в архітектурі ЗНМ (рис. 3.12) залежить від різних чинників:

- *Складність задачі*

Якщо задача вимагає врахування складних і дрібнозернистих особливостей, то більш глибока мережа з більшою кількістю згорткових шарів може бути корисною. Складні завдання, наприклад, виявлення об'єктів або сегментація, часто виграють від глибшої архітектури.

- *Доступність даних*

Розмір і різноманітність доступних даних також можуть впливати на кількість шарів. Більші набори даних зазвичай дають нагоду створювати глибші архітектури, оскільки вони надають більше прикладів, на яких мережа може вчитися. Однак, маючи обмежені дані, важливо уникати надмірного пристосування, яке може статися з надто складними моделями.

- *Обчислювальні ресурси*

Кількість шарів має відповідати обчислювальним обмеженням апаратного забезпечення, що використовується. Більш глибокі архітектури вимагають більше пам'яті та обчислювальної потужності для навчання. Важливо досягти балансу між складністю моделі та доступними ресурсами.

- *Валідація та експерименти*

Часто необхідно проводити валідаційні експерименти, щоб оцінити продуктивність мережевих архітектур із різною кількістю шарів. Порівнюючи точність і продуктивність узагальнення на валідаційному наборі даних, можна визначити оптимальну кількість рівнів, яка забезпечує хороший компроміс між продуктивністю і складністю.

Найбільш популярним підходом є поступове зменшення просторових розмірів за допомогою застосування максимального об'єднання після кількох шарів згортки. Це допомагає охопити та узагальнити локальні особливості водночас зберігаючи керований розмір карт об'єктів. Конкретна кількість пар шарів згортки та максимального об'єднання зазвичай визначається балансуванням між необхідністю виділення об'єктів та зменшенням розмірності й ризиком втрати інформації та надмірного спрощення. На практиці рекомендується експериментувати з різними конфігураціями архітектури — починати з порівняно простої ($k = 3$) і поступово збільшувати її глибину через додавання шарів згортки та максимального об'єднання. Продуктивність мережі можна відстежувати на валідаційному наборі й зупинити збільшення глибини, коли продуктивність насититься або почне знижуватися.

Верхню межу кількості пар шарів визначає параметр *timePoolSize* функції створення останнього шара пулинга *maxPooling2dLayer*. Змінна *timePoolSize* визначає розмір вікна об'єднання вздовж часового виміру. Вона вказує, скільки суміжних часових кроків об'єднується або агрегується в одне вихідне значення і обчислюється, як ділення кількості часових інтервалів у вхідній спектрограмі *numHops* на 10. Цей розрахунок виконується для того, щоби переконатися, що зона об'єднання часових інтервалів менша за вхідну спектрограму, і щоби уникнути потенційної невідповідності розмірів під час навчання. Велике значення змінної *timePoolSize* призведе до більш агресивного даунсемплінгу, зменшуючи часову роздільну здатність вхідних даних. Це може бути корисно, коли часова динаміка даних не є критично важливою або під час роботи з довгими послідовностями для зменшення обчислювальних вимог. З іншого боку,

менше значення *timePoolSize* зберігає більше часової інформації і може бути корисним, коли для завдання важливі дрібнозернисті часові патерни.

Отже, за результатами розрахунку розмірності вихідних даних розроблювальної моделі архітектури нейронної мережі, було визначено максимальне значення оптимальної кількості пар шарів згортки та максимального об'єднання ($k = 10$), за якого розмір вихідних даних згорткового шару $convolutionalOutputDataSize \geq timePoolSize$.

<u>Шар згладжування</u> (*Flatten Layer*) використаємо для проведення частотно-часової репрезентації вихідних даних з останнього шару об'єднання в одновимірний (згладжений) вектор. Він ефективно «вирівнює» просторову структуру, готуючи дані до подальшої обробки наступними повнозв'язними шарами, які потребують одновимірного входу.

Важливою перевагою застосування шарів згладжування є збереження інформації про вихідні просторові виміри. Це дає змогу наступним шарам мати доступ до просторових відношень між об'єктами. Також цей шар зменшує розмірність вхідних даних, не втрачаючи за такої умови жодної інформації. Він конденсує просторові розміри в один вектор, зменшуючи кількість параметрів, необхідних у наступних шарах. Це може допомогти зменшити надмірну підгонку й підвищити обчислювальну ефективність мережі.

<u>Шар відсікання</u> (Dropout Layer) — це техніка регуляризації, яка використовується для запобігання перенавчанню НМ. Під час навчання шар випадковим чином встановлює частину вхідних одиниць на нуль за кожного оновлення, фактично "відсікаючи" ці одиниці. Це допомагає запобігти сумісній адаптації нейронів і змушує мережу покладатися на підмножину одиниць для прогнозування, у такий спосіб, запобігаючи надмірній залежності мережі від специфічних особливостей або запам'ятовуванню навчальних даних.

У програмі Матлаб шар відсікання реалізовується наступною командою:

```
dropoutLayer(probability)
```

де «*probability*» — імовірність, що поточний вхідний елемент буде обнулений.

У програмі значення змінної «*robability*» встановлено у 0.2, тому шар випадковим чином відкине 20% вхідних одиниць даних.

Важливо зазначити, що під час фази оцінювання або висновку, шар відсікання зазвичай вимикається, а для прогнозування використовується вся мережа. Оптимальна ймовірність відсіву може змінюватися залежно від поставленого завдання, набору даних та архітектури мережі, тому, щоб знайти оптимальне її значення для конкретної задачі, необхідно експериментувати.

Наступним кроком під час побудови архітектури ЗНМ для завданнь класифікації є створення **повнозв'язного шару** (*Fully Connected layer*), де кожен нейрон з'єднаний із кожним нейроном попереднього шару. Він помножує вхідні елементи на вагову матрицю, а потім додає вектор зміщення.

У програмі Матлаб цей шар задається наступним чином:

```
fullyConnectedLayer(numClasses),
```

де *numClasses* — кількість вихідних класів або категорій й обчислюється за допомогою підрахунку числа унікальних категорій у масиві даних навчальної вибірки.

Повністю зв'язаний шар виконує лінійне перетворення вхідних даних, після чого застосовується активаційна функція для введення нелінійності. Це дає змогу мережі вивчати складні взаємозв'язки між вхідними характеристиками та цільовими класами. Найпоширенішими функціями активації для повністю зв'язаних шарів є *ReLU* та *Softmax*.

Важливо зазначити, що кількість нейронів у повнозв'язному шарі може впливати на пропускну здатність і складність мережі, тому її потрібно коригувати залежно від конкретного завдання та набору даних.

Використовуючи **шар Softmax** нормалізуємо вихідні значення повнозв’язного шару для отримання прогнозованих ймовірностей, що вказують на ймовірність належності заданих вхідних даних до кожного класу. Функція Softmax підносить вхідні значення до степені експоненти й нормалізує їх за сумою всіх експоненційованих значень. Це гарантує, що результуючі значення будуть додатними й у сумі дорівнюватимуть 1, представляючи собою

ймовірності. Клас із найвищою ймовірністю зазвичай обирається, як передбачуваний клас для вхідних даних.

Варто зазначити, що функція Softmax чутлива до екстремальних значень, що може призвести до чисельної нестабільності. Для вирішення цієї проблеми на практиці часто застосовують такі методи, як функція Log-Sum-Exp — послідовне застосування до вхідного аргументу експоненти, сумування та логарифмування.

Отримане значення максимальної ймовірності пропускаємо крізь **шар класифікації** (*Classification layer*), який приписує максимальне значення ймовірності відповідній мітці класу для заданих вхідних даних на основі активації попередніх шарів. У такий спосіб на виході ми отримуємо мітку класу, а не ймовірність. У програмі Matlab шар класифікації є попередньо визначеним останнім рівнем мережевої архітектури й задається наступним чином:

```
classificationLayer(Classes=classes,ClassWeights=classWeights)];
```

Аргумент «*Classes*» визначає класи або категорії, які мережа має класифікувати. Змінна *classes* представляє масив рядків у наборі даних або категорійний масив, що містить мітки класів. Аргумент «*Classes»* гарантує, що вихідний шар мережі сумісний із кількістю та порядком класів у наборі даних.

Другий аргумент функції — «*ClassWeights*», дає змогу призначати різні ваги кожному класу під час процесу навчання. Він використовується для обробки дисбалансу класів, коли певні класи мають більше або менше зразків, ніж інші. Змінна *classWeights* задає вагу, призначену кожному класу й обчислюється, як взята поелементно обернена величина до кількості входжень кожного класу в навчальній вибірці. Далі ваги класів масштабуються за їхнім середнім значенням, що нормалізує їх так, щоб сума дорівнювала кількості класів.

Шар *classificationLayer* зазвичай використовується разом із функцією активації *softmax* і функцією *перехресних ентропійних втрат*, які добре підходять для завдань багатокласової класифікації. Функція softmax гарантує, що вихідні значення представляють ймовірності класів, тоді, як функція перехресних ентропійних втрат вимірює розбіжність між прогнозованими ймовірностями і справжніми мітками класів. Під час навчання мережа

намагається мінімізувати втрати, налаштовуючи свої параметри для підвищення точності прогнозів.

Повна архітектура спроектованої згорткової нейронної мережі для проведення експериментальних досліджень роботи алгоритму виявлення та класифікації аудіосигналів МЛА, зображена на рисунку 3.13.

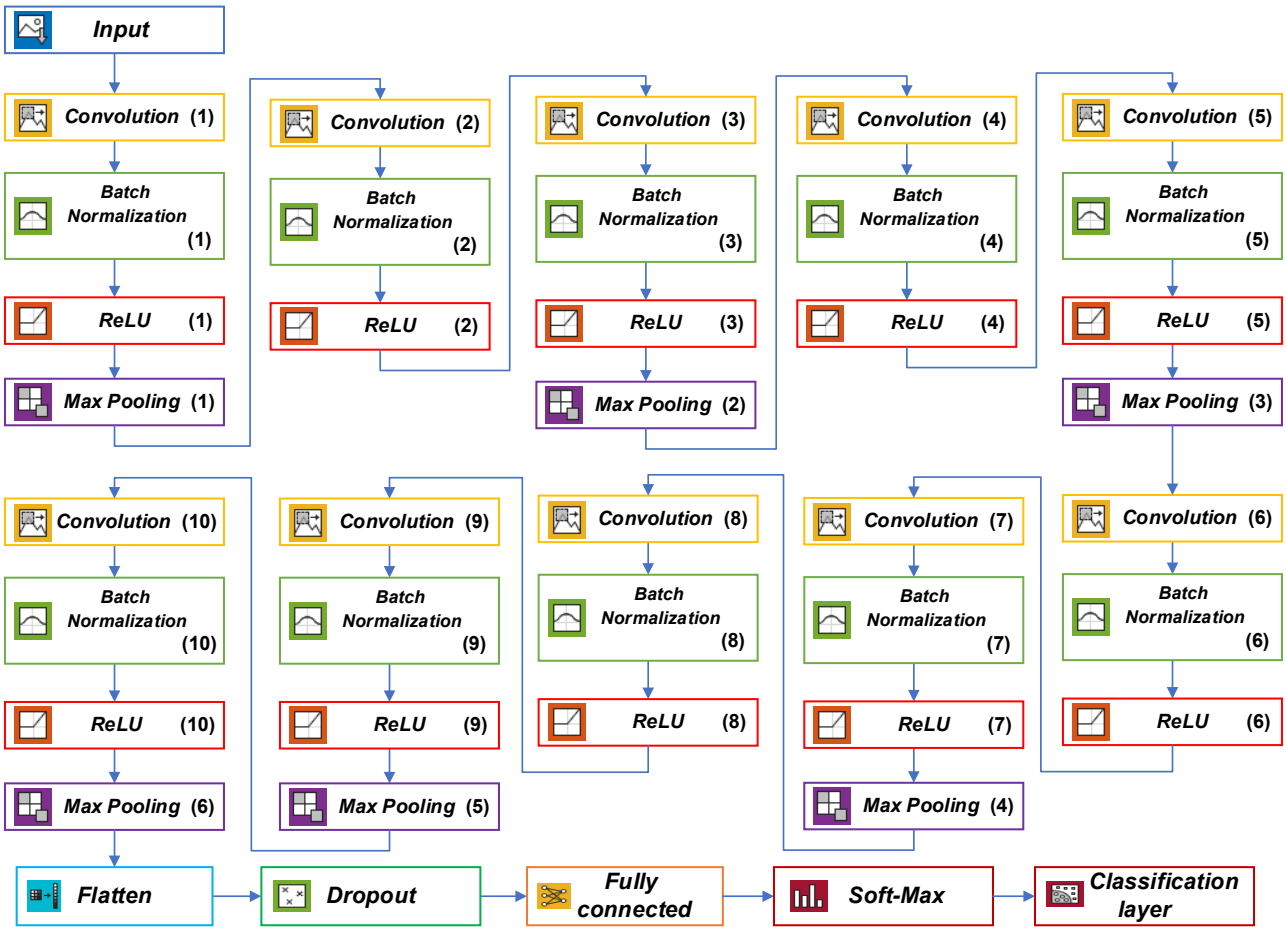


Рисунок 3.13 — Повна архітектура розробленої ЗНМ для класифікації аудіосигналів МЛА

3.5.2. Навчання згорткової нейронної мережі

Процес навчання згорткової нейронної мережі включає в себе подачу виділених ознак аудіозразків навчальних даних у модель, застосування оптимізаторів для оновлення значення коефіцієнтів ваги через зворотнє розповсюдження та обчислення зміщення мережі для мінімізації функції втрат.

Розглянемо найбільш поширені оптимізатори в багатьох додатках глибокого навчання, що довели свою ефективність під час класифікації аудіо.

Стохастичний градієнтний спуск (Stochastic Gradient Descent — *SGD*) оновлює параметри мережі (ваги та зміщення) для мінімізації функції втрат, роблячи невеликі кроки на кожній ітерації в напрямку від'ємного градієнта втрат.

$$\theta_{\ell+1} = \theta_\ell - \alpha \nabla E(\theta_\ell)$$

де ℓ – номер ітерації;

$\alpha > 0$ – швидкість навчання;

θ – вектор параметрів;

$E(\theta)$ – функція втрат.

У стандартному алгоритмі градієнтного спуску, градієнт функції втрат $\nabla E(\theta)$, оцінюється на всій навчальній вибірці, також алгоритм використовує весь набір даних одразу. На відміну від цього, алгоритм стохастичного градієнтного спуску оцінює градієнт і оновлює параметри, використовуючи підмножину навчальних даних. На кожній ітерації використовується інша підмножина, яка називається мініпакетом. Повний прохід алгоритму навчання по всій навчальній множині з використанням мініпакетів становить одну епоху. Стохастичний градієнтний спуск простий та ефективний у обчисленні, тому що може уникати локальних мінімумів завдяки своїй стохастичній природі.

Недоліком цього алгоритму є тенденція до повільної збіжності, що вимагає надто ретельного налаштування швидкості навчання та додавання імпульсу до оновлення параметрів для зменшення коливань вздовж шляху найкрутішого спуску до оптимуму [48] .

Стохастичний градієнтний спуск з імпульсом (Stochastic Gradient Descent with Momentum — *SGDM*) визначається наступною формулою:

$$\theta_{\ell+1} = \theta_\ell - \alpha \nabla E(\theta_\ell) + \gamma(\theta_\ell - \theta_{\ell-1})$$

де γ визначає внесок попереднього кроку градієнта в поточну ітерацію.

Стохастичний градієнтний спуск з імпульсом використовує єдину швидкість навчання для всіх параметрів. Інші ж алгоритми оптимізації намагаються покращити навчання мережі, використовуючи швидкості навчання, які відрізняються за параметрами й можуть автоматично адаптуватися до функції втрат, що оптимізується. **RMSProp** (Root Mean Square Propagation —

середньоквадратичне поширення) є одним із таких алгоритмів. Він зберігає значення ковзного середнього з поелементних квадратів градієнтів параметрів:

$$\vartheta_\ell = \beta_2 \vartheta_{\ell-1} + (1 - \beta_2) \cdot |\nabla E(\theta_\ell)|^2$$

де β_2 – швидкість спадання ковзного середнього (найчастіше 0.9, 0.99 та 0.999).

Відповідна довжина усереднення квадратів градієнтів дорівнює $1 / (1 - \beta_2)$, тобто 10, 100 і 1000 оновлень параметрів відповідно.

Алгоритм RMSProp використовує це ковзне середнє для нормалізації оновлень кожного параметра окремо, коли поділ виконується поелементно:

$$\theta_{\ell+1} = \theta_\ell - \frac{\alpha \nabla E(\theta_\ell)}{\sqrt{\vartheta_\ell + \varepsilon}}$$

де ε – невелика константа, що додається для уникнення ділення на нуль.

Використання RMSProp ефективно зменшує швидкість навчання параметрів із більшими градієнтами і відповідно збільшує швидкість для параметрів із малими градієнтами, що допомагає функції швидше збігатися в неопуклих задачах оптимізації. Але «побічним ефектом» стають проблеми з різною швидкістю навчання для різних параметрів, що вимагає ретельного налаштування гіперпараметрів.

AdaGrad (*Адаптивний градієнт*) — оптимізатор, який дає змогу оновлювати параметри з огляду на те, наскільки типову ознаку вони фіксують. Він, як і RMSProp, адаптує швидкість навчання для кожного параметра залежно від частоти його оновлення, що надає нагоду алгоритму по-різному регулювати кроки оновлення. Також AdaGrad відстежує накопичену суму квадратів градієнтів для кожного параметра та оновлює її на кожній ітерації, додаючи квадрат поточного градієнта. Щоб швидкість навчання не стала занадто малою із часом, AdaGrad також вводить коефіцієнт згладжування. Цей чинник допомагає уникнути ділення на дуже малі значення під час розрахунку швидкості навчання, гарантуючи, що швидкість не впаде до нуля занадто стрімко. Відповідно формула оновлення даних матиме наступний вигляд:

$$\theta_{\ell+1} = \theta_\ell - \frac{\eta}{\sqrt{G_\ell + \varepsilon}} g_t$$

де G_l – сума квадратів оновлень;

ε – згладжуючий параметр, зазвичай порядку 10^{-6} або 10^{-8} для агресивного оновлення. Але, як видно з графіків (рис. 3.14), це відіграє роль тільки на початку, ближче до середини навчання починає переважати G_l :

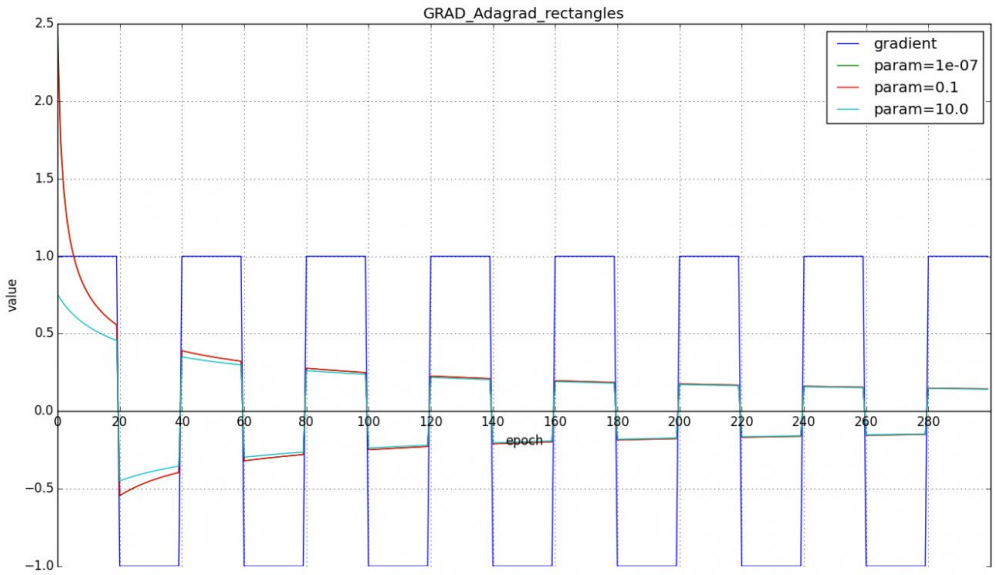


Рисунок 3.14 — Графік зміни параметрів AdaGrad під час навчання

Ще одною перевагою AdaGrad є відсутність необхідності точно підбирати швидкість навчання. Досить виставити її в міру великою, щоб забезпечити хороший запас, але не такою величезною, щоб алгоритм розходився.

Однак, AdaGrad має деякі обмеження: по-перше, накопичення квадратичних градієнтів може призвести до швидкого зменшення швидкості навчання із часом, що приведе до передчасної збіжності й зупинки процесу навчання занадто рано; по-друге, AdaGrad використовує лише інформацію про градієнт першого порядку й не враховує похідні другого порядку, що може обмежити її здатність вирішувати певні оптимізаційні задачі; по-третє, оптимізатор не враховує імпульс, що може сповільнити збіжність у деяких випадках, особливо в задачах із більш складним ландшафтом втрат.

Алгоритм оптимізації **Adadelta**, подібно до Adagrad, накопичує суму квадратів градієнтів із часом для кожного параметра. Вона вводить додаткове ковзне середнє значення попередніх оновлень для коригування швидкості навчання. Це середнє слугує оцінкою моментів градієнтів другого порядку.

Швидкість навчання для кожного параметра визначається на основі накопичених градієнтів і поточного середнього значення попередніх оновлень. Адаптуючи швидкість навчання на основі локальної історії градієнтів, Adadelta долає проблему зменшення швидкості навчання, яка виникає в Adagrad.

Крок оновлення для кожного параметра обчислюється за допомогою ділення середньоквадратичного значення накопичених градієнтів на середньоквадратичне значення попередніх оновлень. Ця нормалізація допомагає стабілізувати оновлення і запобігає різким змінам значень параметрів:

$$\theta_{\ell+1} = \theta_{\ell} - \frac{RMS(\Delta\theta)_{\ell-1}}{RMS(g)_{\ell}} g_t$$

де $RMS(\Delta\theta)_{\ell} = \sqrt{E(\Delta\theta^2)_{\ell} + \varepsilon}$ — стабілізуючий член, пропорційний *RMS* від $\Delta\theta_l$.

У цьому випадку оновлення параметрів відбувається в три етапи: спочатку накопичуємо квадрат градієнта, потім оновлюємо θ , після чого вже $RMS(\Delta\theta)_l$.

Adadelta, як і RMSProp, адаптує швидкість навчання на основі локальної історії градієнта, що дає змогу ефективно обробляти різні величини градієнтів. Це допомагає подолати проблему зменшення швидкості навчання і забезпечує стабільне та ефективне оновлення параметрів. Також вона автоматично регулює швидкість навчання без необхідності ручного налаштування, що робить його зручним вибором для навчання глибоких нейронних мереж.

Однак, Adadelta також має деякі недоліки: введення додаткових гіперпараметрів, таких, як коефіцієнт розпаду та значення ε, які потрібно правильно налаштувати для оптимальної роботи. Adadelta не включає імпульс, що може обмежити її здатність вирішувати певні оптимізаційні задачі й може призвести до повільнішої збіжності.

Оптимізатор ***Adam*** — це адаптивний алгоритм оптимізації швидкості навчання, який використовує оновлення параметрів, подібне до AdaGrad та RMSProp, але з додаванням імпульсу. Adam адаптує швидкість навчання для кожного параметра на основі оцінювання моментів першого порядку (середнє значення) і моментів другого порядку (дисперсія) градієнтів. Це дає змогу

автоматично регулювати швидкість навчання, що може допомогти в досягненні швидшої збіжності:

$$m_\ell = \beta_1 m_{\ell-1} + (1-\beta_1)\cdot \nabla E(\theta_\ell)$$
$$\vartheta_\ell = \beta_2 \vartheta_{\ell-1} + (1-\beta_2)\cdot |\nabla E(\theta_\ell)|^2$$

де β_1 і β_2 – швидкості спаду.

Adam використовує ковзне середнє для ефективного оновлення параметрів мережі у вигляді:

$$\theta_{\ell+1} = \theta_\ell - \frac{\alpha m_\ell}{\sqrt{\vartheta_\ell + \varepsilon}}$$

Якщо градієнти на багатьох ітераціях схожі, то використання ковзного середнього градієнта дає змогу оновленню параметрів набирати оберти в певному напрямку. Якщо ж градієнти містять переважно шум, то ковзне середнє градієнта стає меншим, отже, й оновлення параметрів також стають меншими. Це корисно в сценаріях, де деякі градієнти можуть бути розрідженими або мати більшу дисперсію. Повне оновлення функції Adam включає механізм для виправлення упередження, яке з'являється на початку навчання [49].

Основними перевагами оптимізатора Adam є поєднання сильних сторін AdaGrad та RMSProp, також він найкраще справляється з різноманітними наборами даних і завданнями, включно із класифікацією аудіо. Він ефективний у неопуклих задачах оптимізації і може збігатися швидше, порівнюючи з традиційним стохастичним градієнтним спуском (SGD). З недоліків можна виділити високу чутливість до гіперпараметрів, які потрібно ретельно налаштовувати для оптимальної роботи. У деяких випадках оптимізатор Adam може демонструвати надмірне налаштування, особливо на невеликих наборах даних. Для пом'якшення цієї проблеми можна використовувати методи регуляризації, такі, як відсіювання або зменшення ваги.

Отже, виберемо найкращий оптимізатор для алгоритму навчання НМ. Оскільки на вхід мережі подається великий набір даних аудіосигналів МЛА, тоді рекомендується використовувати такі оптимізатори, як Adam або RMSProp. Ці оптимізатори адаптують швидкість навчання для кожного параметра на основі

величини останніх градієнтів, що допомагає швидше збігатися і переміщатися через плоскі ділянки ландшафту втрат. Це робить їх найбільш придатними для великомасштабних сценаріїв навчання. Але Adam має перевагу в цьому аспекті, оскільки використовує оцінки не тільки першого, а і другого моменту для нормалізації швидкості навчання. Також Adam добре працює з наборами даних, які мають рідкісні градієнти, що часто трапляється в аудіоданих. Він ефективно адаптує швидкість навчання до різної величини градієнта, що дає змогу моделі прогресувати, навіть за рідкісних його оновлень. Але зі збільшення вхідних даних та аналітичних можливостей функцій навчання НМ, зростає і рівень необхідної потужності та об’єму пам’яті обчислювального пристрою. За умови вибору оптимізаторів, таких як Adam та RMSProp, необхідно мати потужний ПК, тому що в задачі виявлення та класифікації дронів вкрай важлива не тільки ефективність, а і швидкість роботи алгоритму.

Зрештою, для поставленого завдання найкращим оптимізатором є **Adam**, оскільки він видає хороші результати вже за мінімального підгону параметрів. Але продуктивність будь-якого оптимізатора сильно залежить від правильно підібраних ключових гіперпараметрів: розміру партії, швидкості навчання, кількості ітерацій та епох, імпульсу та розподілу ваги (якщо застосовуються). Тож розрахуємо їхні значення для отримання максимальної продуктивності.

Мініпартія — це підмножина навчальної вибірки, яка обробляються за кожної ітерації для оцінювання градієнта функції втрат та оновлення ваг моделі. Розмір мініпартії (*mini-batch size*) має бути таким, щоб легко вписувався в обсяг доступної пам'яті системи обчислення. Більший розмір може призвести до швидшого навчання, оскільки паралельно обробляється більше прикладів. Однак такий розмір не завжди призводить до кращих результатів і може сповільнити навчання через обчислювальні обмеження.

Під час навчання на менших наборах даних або роботі з моделями, які схильні до перенавчання, корисно використовувати менші розміри мініпартій. Вони можуть створювати більше шуму в оцінках градієнта, що матиме

регуляризуючий ефект і покращує узагальнення. Однак, дуже малі розміри можуть призвести до нестабільності та повільнішої збіжності.

Також треба враховувати й апаратне забезпечення для навчання мережі. Деякі апаратні прискорювачі, такі, як графічні процесори, краще працюють із певною кількістю мініпартій, тому варто поекспериментувати, щоб знайти оптимальне налаштування для конкретної апаратної конфігурації. На практиці зазвичай починають із розмірів у діапазоні від 32 до 256. Якщо розмір мініпартії не рівномірно ділить кількість навчальних вибірок, то за допомогою функції *trainNetwork* програми Матлаб відкидаємо навчальні дані, які не входять в остаточний повний мініпакет кожної епохи.

За результатом емпіричних досліджень, оптимальне значення розміру мініпартії для завдання класифікації аудіосигналів, використовуючи процесор *Intel Core I7*, дорівнює ***128***, що забезпечує найкращий компроміс між використанням пам'яті ПК, швидкістю навчання та продуктивністю моделі.

Вибір кількості ітерацій та епох для навчання НМ передбачає пошук правильного балансу між збіжністю моделі та обчислювальною ефективністю. *Ітерація* — один крок в алгоритмі градієнтного спуску до мінімізації функції втрат за допомогою мініпартії. Кількість ітерацій дорівнює частоті валідації даних (*Validation frequency*) нейронної мережі і визначається, як загальна кількість навчальних даних поділена на розмір мініпартії.

Кількість ітерацій впливає на час навчання, особливо, якщо робота відбувається з великими наборами даних. Навчання з великою кількістю кроків займає більше часу і вимагає значних обчислювальних витрат, у той час, як навчання з малою кількістю ітерацій може призвести до того, що модель не буде пристосована. Тому важливо збалансувати доступні обчислювальні ресурси та бажаний час навчання.

Процес навчання повторюється протягом декількох *епох*. Епохою називають повний прохід алгоритму навчання через усю навчальну вибірку. Максимальна кількість епох (*Max Epochs*), задається натуральним числом. Якщо модель під час навчання не збігається, потрібна збільшувати їхню кількість.

Однак, якщо модель уже збіглася, навчання на великій кількості епох не дасть значних покращень. Для уникнення цього ефекту, впровадимо ранню зупинку, яка відстежує роботу моделі на валідаційному наборі. Якщо продуктивність мережі не покращується або починає гіршати після певної кількості епох, навчання зупиняється достроково.

Також за надмірного пристосування моделі можна застосувати методи регуляризації. Першим використаємо метод перехресної валідації — розбиття набору даних на декілька підмножин для навчання моделі на різних згинах. Це дає можливість спостерігати середню продуктивність на різних епохах та визначати відповідну їхню кількість для вищого рівня узагальнення. Наступним застосуємо графік *швидкості навчання*. Під час вибору швидкості навчання необхідно враховувати характеристики оптимізатора, який використовується. Adam та RMSProp зазвичай працюють із більшою швидкістю навчання, у той час простіші оптимізатори, наприклад, SGD, потребують меншої.

Відповідно до характеристик набору даних та архітектури мережі, налаштуємо швидкість навчання так, щоби після проходження *75%* епох вона поступово зменшувалась у *10* разів. Це дасть змогу моделі робити менші оновлення в міру наближення до оптимального рішення, покращуючи збіжність. Почнемо із загальноприйнятих значень швидкості навчання *0.1*, *0.01* або *0.001*.

3.6. Оцінка та оптимізація розробленої моделі нейронної мережі

Оцінимо продуктивність роботи моделі нейронної мережі на навчальному (без доповнення даних) та валідаційному наборах. Мережа достатньо точна на цих наборах, однак навчальні, валідаційні та тестові дані мають схожий розподіл, який не обов'язково відображає реальні умови. Це обмеження, особливо стосується категорії unknown, яка містить шуми оточуючого середовища.

На рис. 3.15 показано прогрес навчання розробленої НМ на кожній ітерації. На графіках показані значення основних метрик оцінювання мережі для завдання ефективної класифікації аудіосигналів МЛА:

- *Точність навчання* (*Training accuracy*) — точність класифікації для кожної окремої мініпартії.
- *Згладжена точність навчання* (*Smoothed training accuracy*) отримана за допомогою застосування алгоритму згладжування до точності навчання. Вона менш зашумлена, ніж незгладжена, що полегшує виявлення тенденцій.
- *Точність перевірки* (*Validation accuracy*) — точність класифікації на всій перевірочній вибірці.
- *Втрати* та *згладжені втрати під час навчання*, *втрати під час валідації* (*Training loss, smoothed training loss*, and *validation loss*) — втрати на кожній мініпартії, її згладжена версія та втрати на валідаційній вибірці відповідно. Оскільки кінцевим шаром мережі є шар класифікації, то функцією втрат є перехресна втрата ентропії.

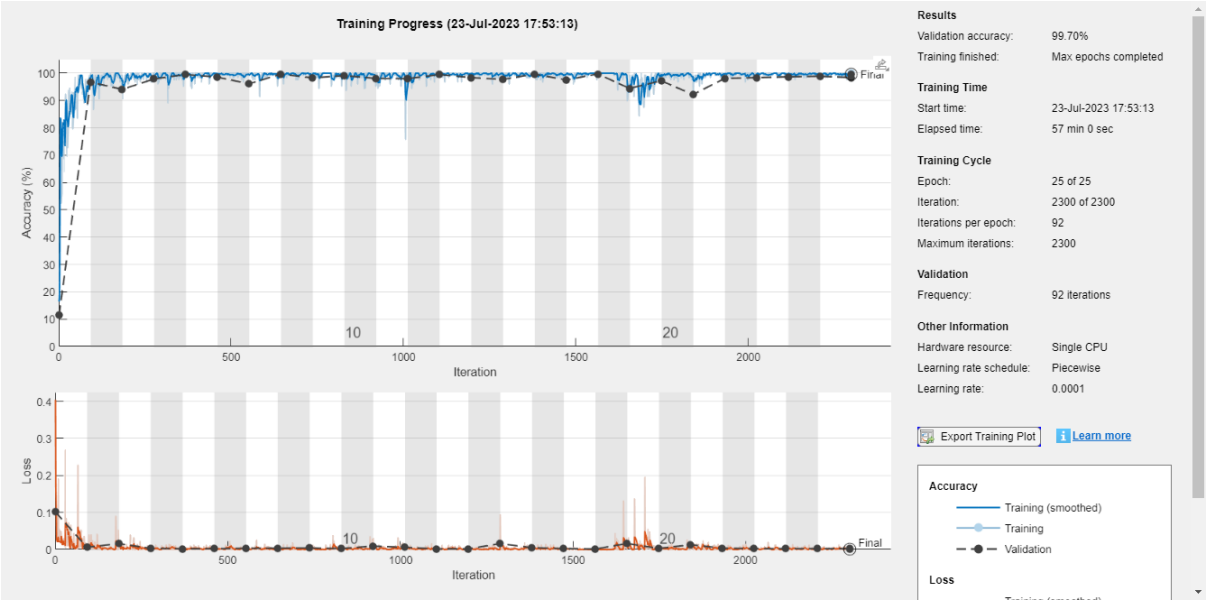


Рисунок 3.15 — Графіки метрик прогресу навчання нейронної мережі

Праворуч міститься основна інформація про час навчання та налаштування мережі. За описаних вище параметрах та архітектурі повний час навчання системи складає 57 хв зі швидкістю 92 ітерації на епоху. Загальна кількість ітерацій дорівнює 2300, а епох — 25.

На основі порівняння передбачених і фактичних міток обчислимо відсоток помилок валідації (*validationError*) та навчання (*trainingError*):

$$validationError = 0.3\ \%\ ; \qquad trainingError = 0.297\ \%\ ;$$

За отриманими даними алгоритму, який передбачає належність кожного об'єкта до одного з класів, побудуємо матрицю помилок (*confusion matrix*) — таблицю, яка використовується для опису продуктивності класифікаційної моделі на наборі даних, для яких відомі істинні значення. Матриця показує кількість правильних і неправильних прогнозів, зроблених моделлю, порівнюючи з фактичними результатами (рис. 3.16). Отримана матриця має розмір 5 на 5 елементів, де значення основної діагональні відповідає правильно визначеним класам із навчальної вибірки, інші ж значення — усім комбінаціям помилкових рішень. Видно, що за отриманої кількості даних, помилка виникає під час розпізнавання моделі *«Feilun Fx137»*.

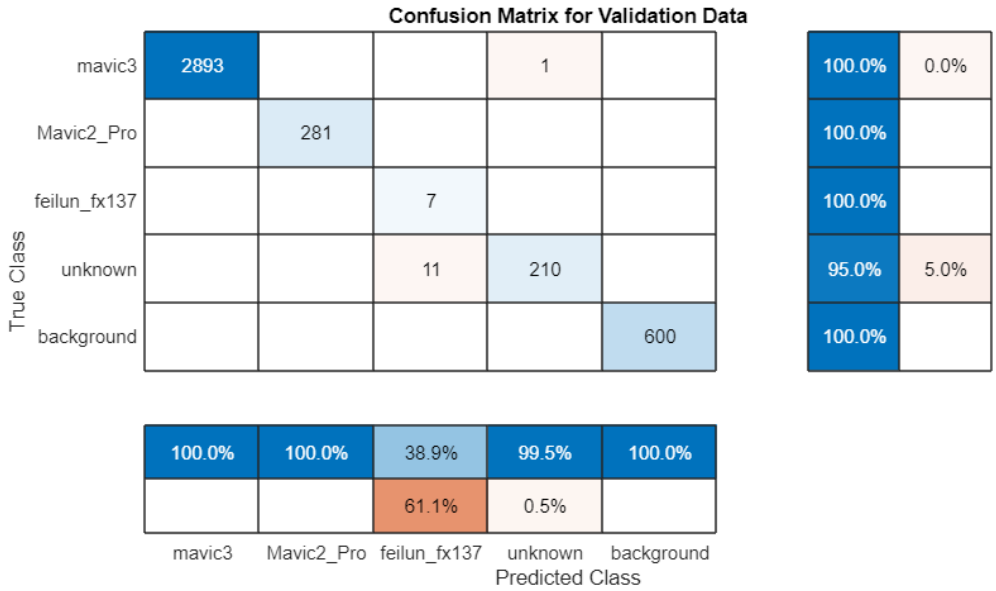


Рисунок 3.16 — Матриця помилок для набору даних валідації

Використовуючи матрицю помилок, визначимо повну оцінку якості роботи алгоритму НМ, для цього розрахуємо основні метрики оцінювання: *precision*, *recall* і *показник F1*, окремо для кожного класу та загальне значення.

Точність (*precision*) — це здатність моделі правильно відрізняти відповідний клас від інших класів. Вона розраховується, як відношення істинно позитивних (*True Positive – TP*) зразків — значення основної діагоналі матриці помилок, до суми істинно позитивних і хибнопозитивних (*False Positive – FP*) зразків — сума значень відповідного рядка матриці:

$$Precision = \frac{TP}{TP + FP} \tag{3.7}$$

Відновлюваність (*recall*), також відома, як чутливість або частота істинних спрацьовувань, вимірює здатність моделі виявляти зразки відповідного класу із загальної кількості фактично позитивних зразків. Вона обчислюється, як відношення кількості істинно позитивних зразків до суми істинно позитивних і хибнонегативних (*False Negative – FN*) зразків — сума значень відповідного стовпця матриці:

$$Recall = \frac{TP}{TP + FN} \tag{3.8}$$

Розраховані за формулами 3.7 та 3.8 значення метрик оцінювання занесемо в таблицю 3.2.

Таблиця 3.2 — «Значення метрик оцінювання ефективності НМ»

Класи	***Mavic 3***	***Mavic 2 Pro***	***Feilun Fx137***	***unknown***	***background***	**ЗАГАЛОМ**
Точність	0,9997	1,0000	1,0000	0,9502	1,0000	0,99
Recall	1,0000	1,0000	0,3889	0,9953	1,0000	0,877

Аналізуючи результати таблиці 3.2 видно, що загальна точність розпізнавання НМ вхідних аудіо даних складає ***99.7 %***, що є відмінним результатом для порівняно невеликої кількості даних.

Оцінка F1 — це середнє гармонійне значення двох метрик: точності та відновлюваності. Вона забезпечує збалансовану метрику оцінювання — враховує помилки типу I (хибнопозитивні) та типу II (хибнонегативні), і розраховується наступним чином:

$$F1 = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{3.9}$$

Показник F1 варіюється від 0 до 1, де значення 1 означає, що модель досягла одночасно високої точності та відновлювання. Він особливо корисний, коли є дисбаланс у розподілі класів, оскільки він рівною мірою враховує як точність, так і відновлювання.

За виразом 3.9 значення показника F1 дорівнює ***0.93***, що вказує на дуже високий рівень загальної продуктивності архітектури мережі з погляду правильної ідентифікації позитивних зразків.

Висновки до розділу 3

1. У розділі 3 було створено алгоритм для виявлення та класифікації аудіосигналів МЛА із використанням методу машинного навчання (рис.3.1).

2. Для підвищення якості набору аудіозаписів дронів та фонових шумів, отриманих під час експерименту в розділі 2, застосували методи попередньої обробки: нормалізацію, фільтрацію, зменшення рівня шуму за допомогою Вінеровської фільтрації, сегментацію (поділили аудіосигнал на менші кадри тривалості *25* мс із перекриттям *50%*) та віконне керування (використовуємо вікно Хеммінга, оскільки в завданні оброблення акустичних сигналів точність у часовій області важливіша).

3. Для представлення даних спрощеним набором акустичних ознак, окремо з кожного кадру попередньо оброблених аудіосигналів, вилучили мел-частотні кепстральні коефіцієнти та визначили мел-спектрограми для фіксації часових та спектральних характеристик. Діапазон частот аналізу знаходиться в межах від *20* до *20* кГц, що становить межі робочих частот обраної моделі мікрофону, а бажана частотна розділова здатність *50* Гц, тому кількість робочих смуг мел дорівнює *30*.

4. Найкращим оптимізатором для ефективного навчання моделі є **Adam**, оскільки він видає хороші результати вже за мінімального підгону параметрів.

5. Для підвищення точності і здатності до узагальнення моделі, застосували методи оптимізації: налаштування гіперпараметрів, швидкості навчання та методів регуляризації.

6. На основі порівняння передбачених і фактичних міток відсоток помилок валідації дорівнює *0.3%,* а навчання — *0.297%.*

7. Основні положення побудови та оцінювання ЗНМ, розглянуті у розділі 3, опубліковані автором у роботі [50].

РОЗДІЛ 4. РЕЗУЛЬТАТИ РОБОТИ СИСТЕМИ ВИЯВЛЕННЯ МАЛИХ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ

4.1. Тестування ефективності роботи системи виявлення МЛА

Після того, як модель навчена й налаштована, протестуємо її роботу на наборі даних, які мережа не бачила в процесі навчання. На першому етапі, у якості тестових даних використаємо аудіозаписи зависання МЛА моделі «*Mavic 3*» на висоті 5 метрів над поверхнею землі та відстані до мікрофону пристрою від 2 до 300 м за рівня фонових шумів оточуючого середовища близькому до нуля. Подамо 10 фрагментів відповідних аудіосигналів на вхід навченої моделі ЗНМ і визначимо ймовірность правильного спрацювання алгоритму (рис. 4.1).

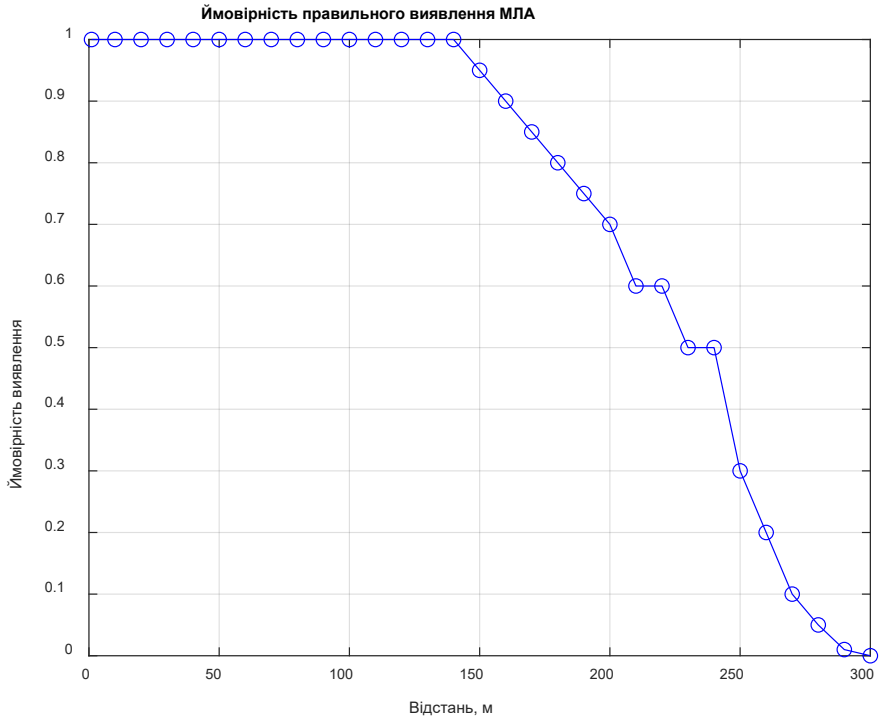


Рисунок 4.1— Ймовірність виявлення МЛА алгоритмом

На графіку видно, що 100% ймовірність виявлення безпілотників відбувається на дистанції до 140 м. Далі вона починає зменшуватися залежно від збільшення відстані до об’єкту.

Створимо критерій для поняття *«ефективне детектування»* — це відстань, на якій безпомилкове виявлення дронів відбувається з ймовірністю не менше 0.7. Відповідно ефективна дистанція роботи алгоритму складає 200 м.

На другому етапі визначення якості роботи алгоритму, використаємо метод експертних оцінок. Оскільки під час роботи над пристроєм у його структурі не використовувалися складні акустичні системи, то в ролі «експертів» будуть виступити 5 людей-операторів, які також будуть визначати наявність МЛА моделі «*Mavic 3*» за таких самих умов, що і НМ. Для зручності, отримані дані експерименту занесемо в таблицю 4.1. Їхній аналіз дасть відповідь на важливе питання: «чи зможе розроблена нейронна мережа замінити оператора в завданні ефективного виявлення малих безпілотних літальних апаратів?» Комірки позначені «+», означають, що експерт чітко чує дрон на відповідній відстані, «+/–» показує, що людина щось чує, але напевно визначити не може, а «–», коли не чутно шум МЛА.

Таблиця 4.1 — Результати виявлення МЛА

Відстань, *м*		70	80	90	100	110	120	130	140	150	160	170	180
Експерти	***1***	+	+	+	+	+ / –	+ / –	–	–	–	–	–	–
	2	+	+	+	+ / –	+ / –	–	–	–	–	–	–	–
	3	+	+	+	+	+	+	+ / –	–	–	–	–	–
	4	+	+	+	+ / –	+ / –	–	–	–	–	–	–	–
	5	+	+	+	+	+	+	–	–	–	–	–	–
ЗНМ		+	+	+	+	+	+	+	+	+	+	+ / –	+ / –

Аналізуючи отримані дані, можна зробити висновок, що розроблений алгоритм, завдяки своєму вражаючому радіусу ефективного виявлення безпілотників у 200 метрів, перевершує діапазон можливостей людського слуху, який обмежений максимальною відстанню в 120 метрів.

Також, на відміну від людини, розроблений алгоритм працює 24/7, не втомлюючись, забезпечуючи постійну пильність та своєчасне виявлення дронів у різних умовах середовища. Він може ефективно обробляти величезні обсяги аудіоданих, швидко розрізняючи звичайні навколишні звуки та виразні акустичні сигнатури МЛА. Крім того, алгоритм усуває потенційну можливість людської помилки, суб'єктивності та обмежень, пов'язаних із людським

сприйняттям. Автоматизуючи процес виявлення, він зменшує ймовірність помилкових спрацьовувань або пропущених виявлень, гарантуючи вищий рівень точності та загальної продуктивності.

Підсумовуючи вище сказане, розроблений алгоритм однозначно може замінити людський персонал в операціях спостереження за малими безпілотними літальними апаратами, пропонуючи підвищену ефективність, надійність і безпеку для громадян, об’єктів критичної інфраструктури та національної оборони.

На фінальному етапі, перевіримо ефективність роботи алгоритму, коли на вхід НМ подаються ті самі фрагменти аудіозаписів зависання МЛА, але з накладанням на них фонових шумів із різним співвідношенням С/Ш від 40 дБ до 12 дБ за потужністю (табл. 4.2).

Таблиця 4.2 — Ймовірність виявлення МЛА за різного значення С/Ш

Відст,м	**С/Ш (дБ)**										
	40	***37***	***35***	***33***	***30***	***27***	***24***	***21***	***18***	***15***	***12***
2	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	0,5
20	1	1	1	1	1	1	1	1	1	1	0,5
25	1	1	1	1	1	1	1	1	1	1	0,3
30	1	1	1	1	1	1	0,7	*0,7	*0,6	0,3	0
35	1	1	1	1	1	1	1	1	1	*0,7	*0,3
40	1	1	1	1	1	0,7	0,7	*0,3	*0,3	*0,3	*0,3
45	0,7	0,7	0,6	0,5	0,6	0,5	0,4	*0,2	*0,3	*0,2	*0,4
50	0,8	0,8	0,7	0,6	0,4	0,4	0,2	*0,2	*0,3	*0,2	*0,7
60	0,5	0,5	0,3	0,4	0,2	0,3	0,2	*0,2	*0,2	*0,2	*0,3
70	0,5	0,4	0,3	0,3	0,2	0,3	0,1	*0,2	*0,!	0	*0,2
80	0,5	0,4	0,4	0,4	0,2	0,2	0,2	*0,1	*0,2	*0,3	*0,3
90	0,3	0,3	0,2	0,2	0,2	0,1	0,2	*0,1	*0,3	0	0
100	0,3	0,2	0,2	0,2	0,2	0,1	0,2	*0,1	0	0	0
110	0,3	0,2	0,2	0,2	0,3	0,2	0,1	*0,1	0	0	0
120	0,3	0,2	0,1	0,2	0,2	0,1	0	0	0	0	0
130	0,3	0,2	0,1	0,2	0	0	0	0	0	0	0
140	0,2	0,1	0	0,2	0	0	0	0	0	0	0
150	0,2	0,1	0	0	0	0	0	0	0	0	0

Примітка. * позначені значенням ймовірностей у тих комірках, де мережа виявила об’єкт, але помилилась із визначенням моделі беспілотника.

За даними табл. 4.2 побудуємо графіки залежності ймовірності виявлення об’єкту від відстані (рис. 4.2 – 4.13).

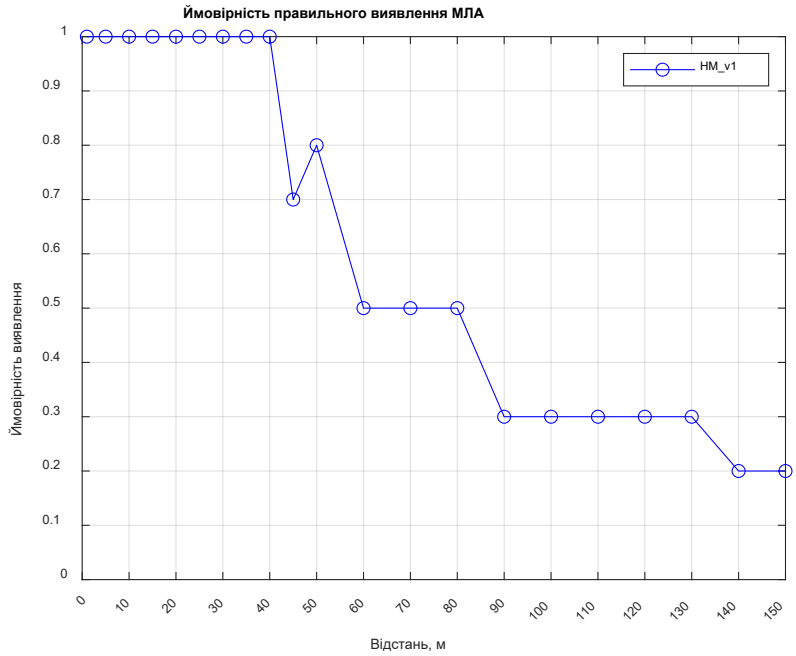


Рисунок 4.2 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 40 дБ

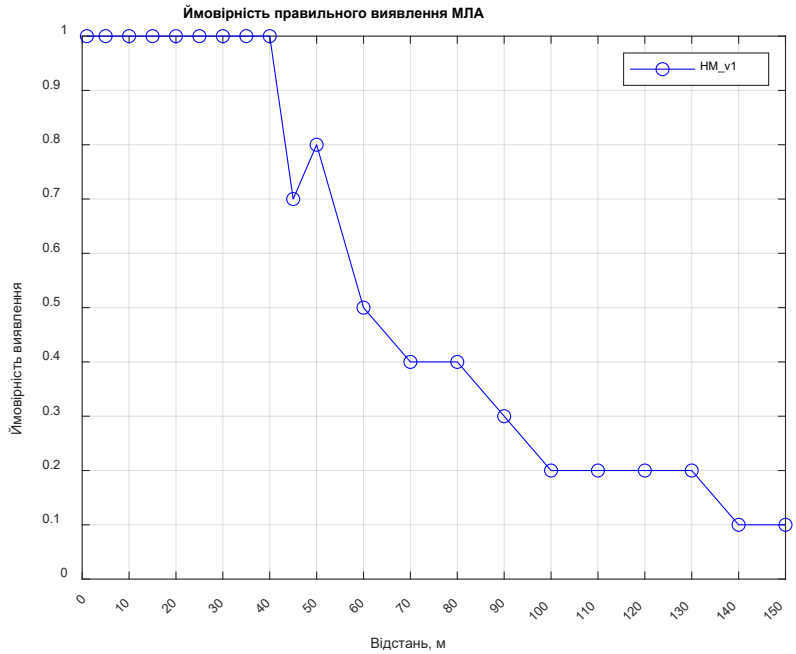


Рисунок 4.3 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 37 дБ

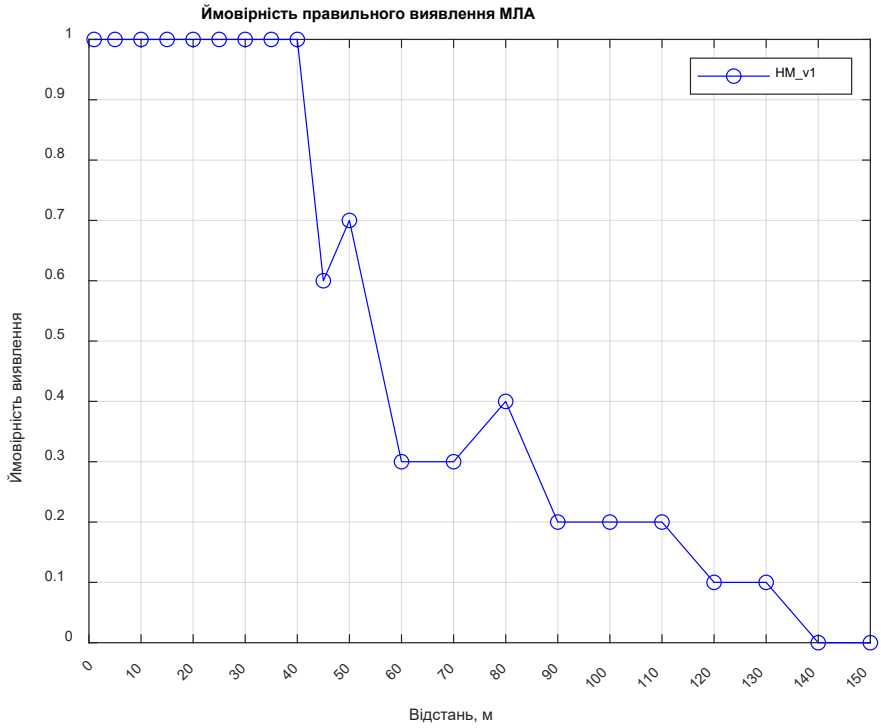


Рисунок 4.4 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 35 дБ

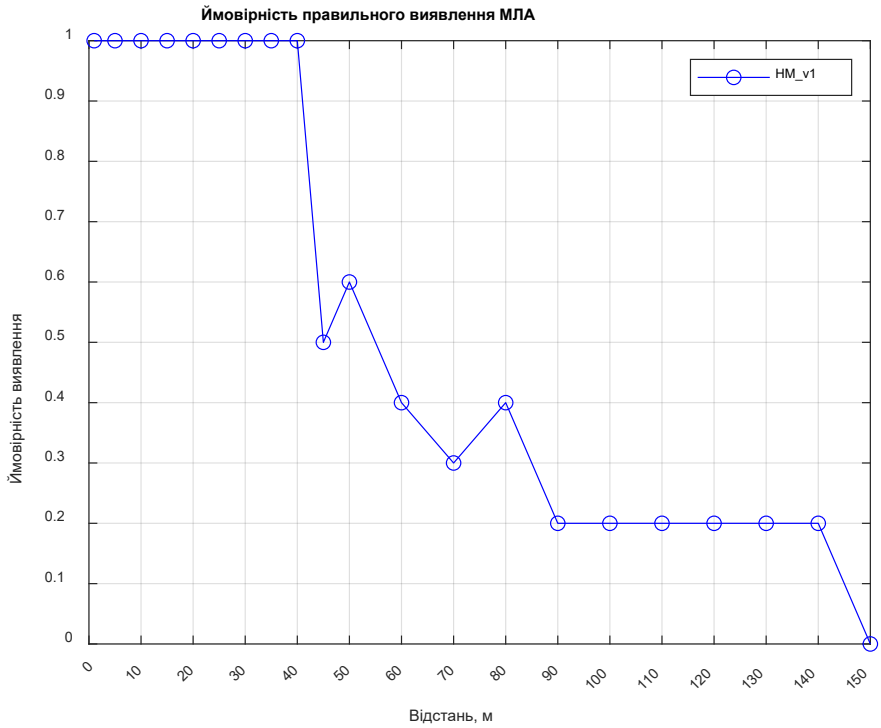


Рисунок 4.5 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 33 дБ

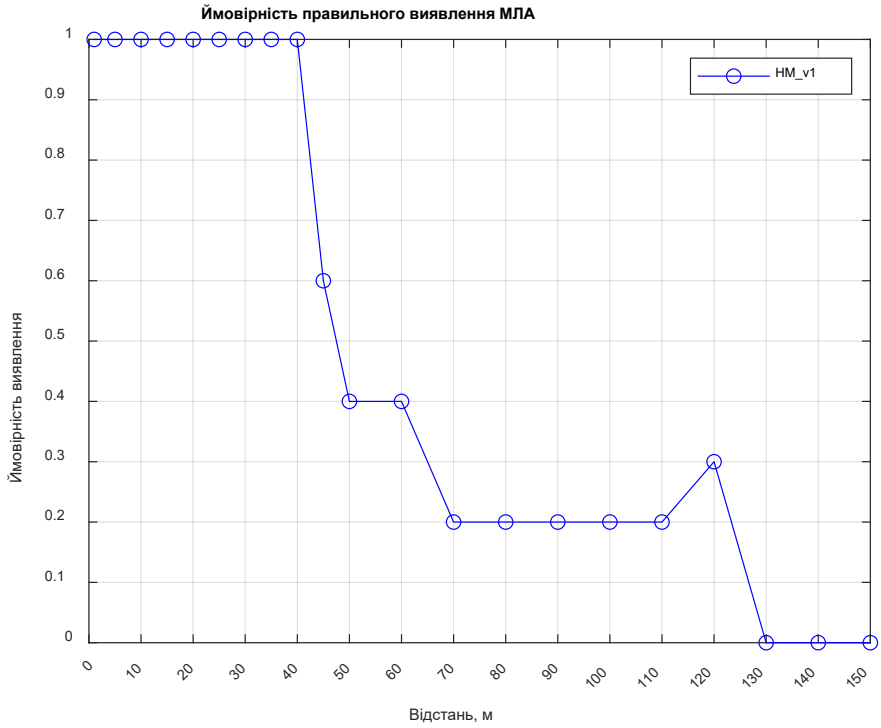


Рисунок 4.6 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 30 дБ

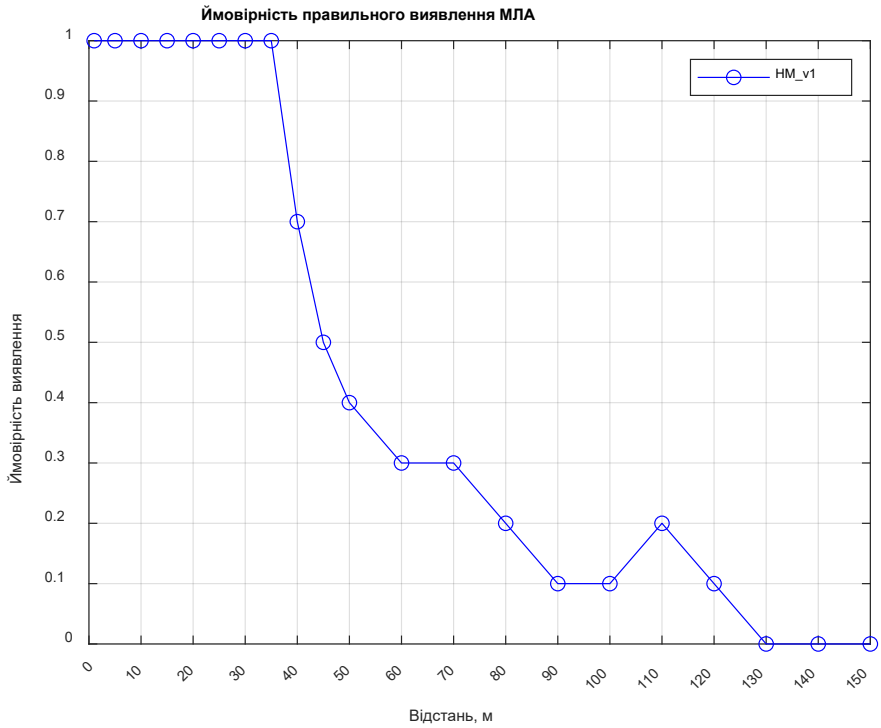


Рисунок 4.7 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 27 дБ

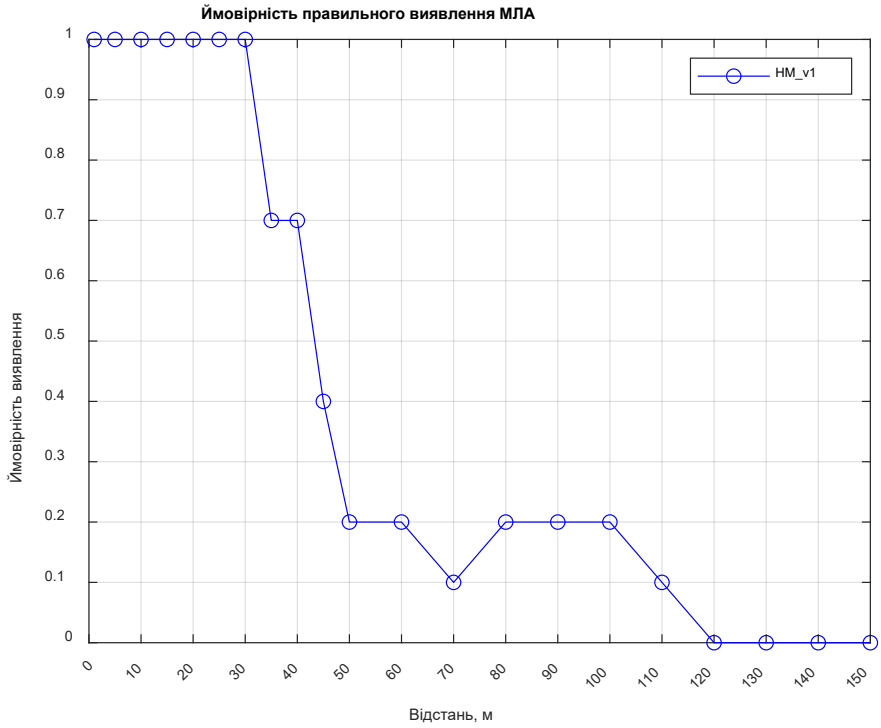


Рисунок 4.8 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 24 дБ

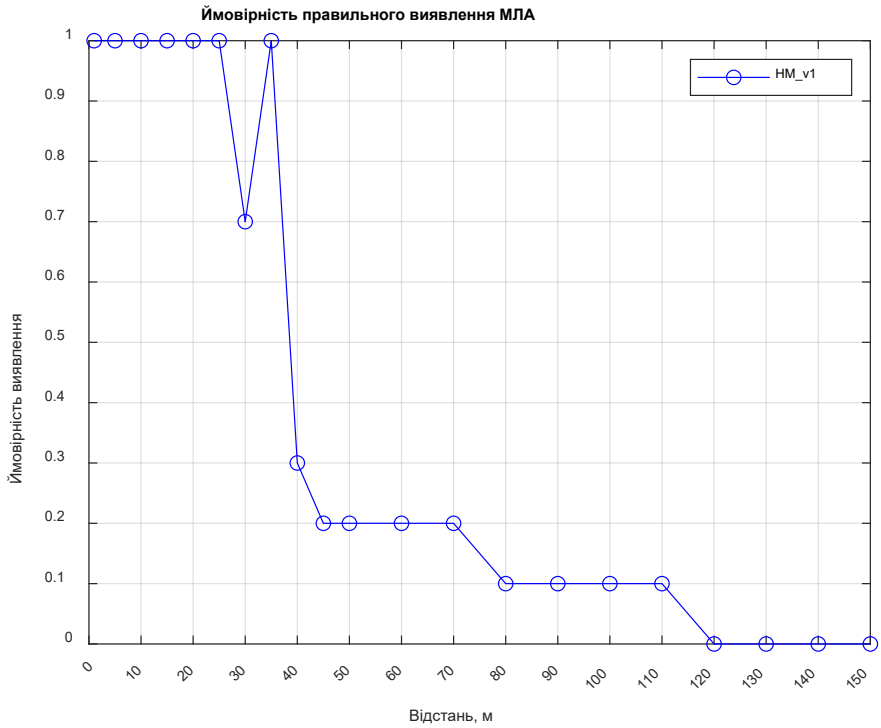


Рисунок 4.9 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 21 дБ

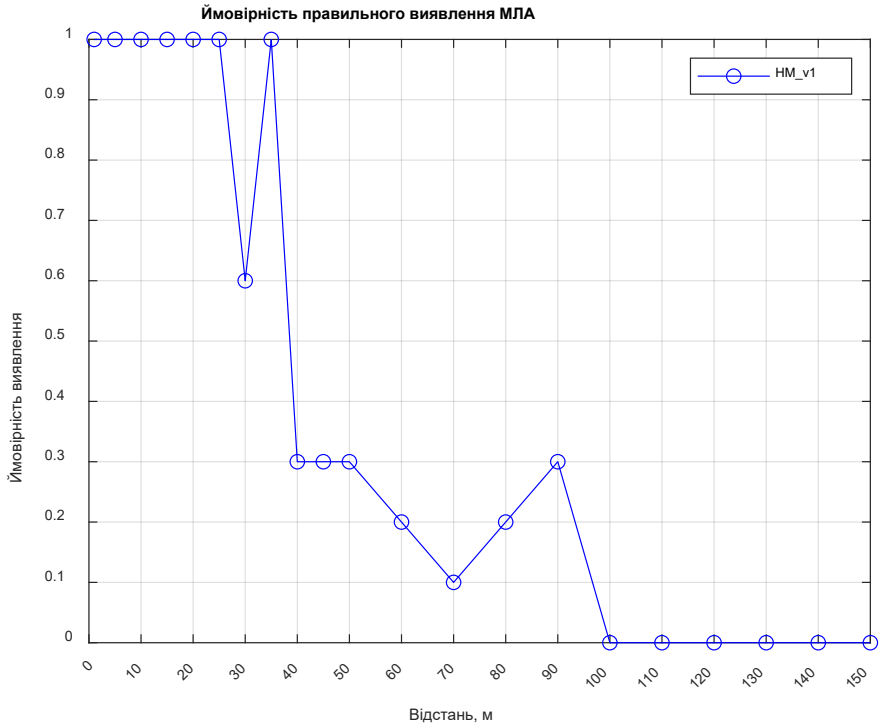


Рисунок 4.10 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 18 дБ

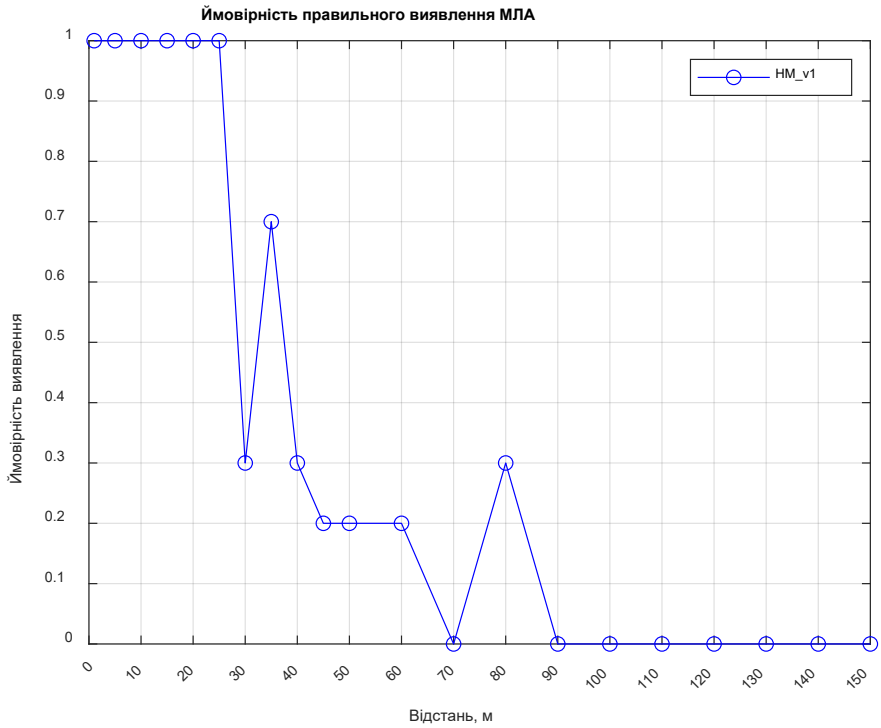


Рисунок 4.11 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 15 дБ

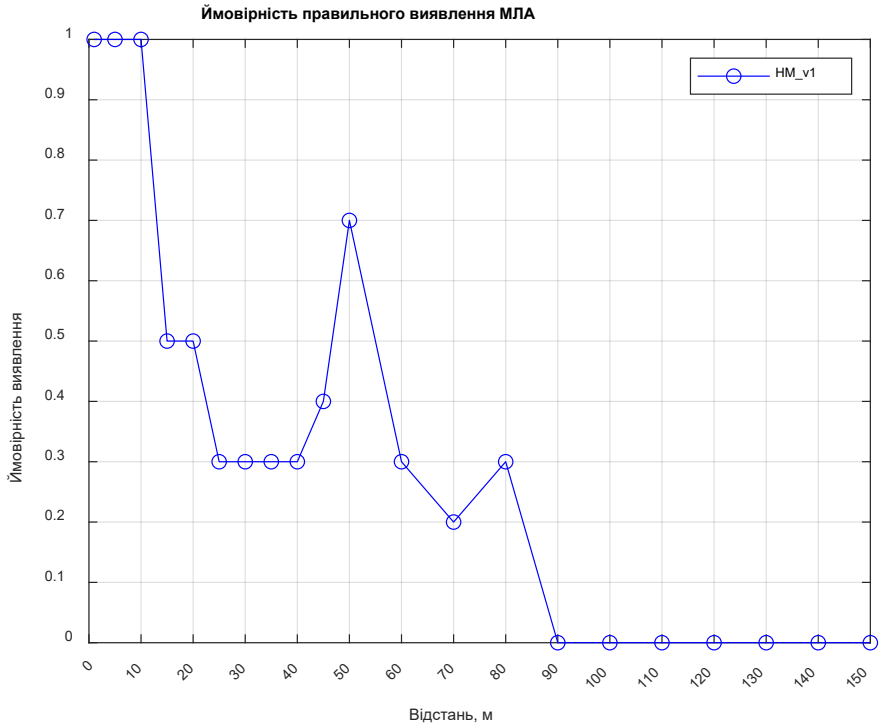


Рисунок 4.12 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 12 дБ

Проаналізувавши експериментальні дані, робимо висновок, що зі збільшенням рівня фонових шумів, тобто зменшенням співвідношення С/Ш, точність виявлення МЛА спадає залежно від росту відстані до об’єкту.

Для подальшого підвищення ефективності розпізнавання сигналів дронів, найкращим підходом є збільшення та покращення набору даних навчання НМ: збір аудіозразків різних моделей безпілотників, включно зі схемами польотів на відповідних дистанціях та висотах, що визначаються можливостями апаратного забезпечення (розділ 2), а також за різного рівня шумів навколишнього середовища. Це дає змогу алгоритму ефективніше навчатися та узагальнювати дані.

Перевіримо це твердження на практиці. Для збільшення різноманітності бази аудіосигналів, скористаємось методом *аугментації*, деформувавши цільові аудіозаписи зависання МЛА за допомогою додавання адитивного білого гаусівського шуму зі співвідношенням С/Ш від 40 дБ до 27 дБ за потужністю. Проведемо навчання НМ використовуючи оновлений *AudioDatastore*.

Для більш ретельної перевірки ефективності програми, на вхід модернізованої нейронної мережі (далі *НМ_v2*) подамо 20 тестових сигналів цілі

з різним співвідношенням С/Ш від 40 дБ до 12 дБ за потужністю, які НМ не бачила в процесі навчання (табл. 4.3).

Таблиця 4.3 — Ймовірність виявлення *НМ_v2* за різного значення С/Ш

Відст,м	**С/Ш (дБ)**										
	40	***37***	***35***	***33***	***30***	***27***	***24***	***21***	***18***	***15***	***12***
2	1	1	1	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1	1
15	1	1	1	1	1	1	1	1	1	1	0,5
20	1	1	1	1	1	1	1	1	1	1	0,5
25	1	1	1	1	1	1	1	1	1	0,6	0,4
30	1	1	1	1	1	1	1	1	1	0,3	0,45
35	1	1	1	1	1	1	1	1	0,75	0,3	0
40	1	1	1	1	1	1	1	1	0,75	0,3	0,15
45	1	1	1	1	1	1	1	1	1	0,1	0
50	1	1	1	1	1	1	1	1	0,7	0,1	0
60	0,8	0,9	0,85	0,9	1	1	1	0,95	0,85	0,15	0
70	0,75	0,85	0,9	1	1	1	1	1	0,8	0	0,1
80	0,7	0,8	0,9	1	1	1	0,8	0,85	0,9	0,15	0
90	0,75	0,8	0,85	1	1	1	1	1	0,75	0	0
100	0,7	0,7	0,75	0,9	1	1	1	0,8	0,85	0	0,15
110	*0,7	*0,7	0,7	0,7	1	0,95	1	1	0,85	0	0
120	*0,7	*0,7	0,65	0,65	0,95	1	1	0,85	0,75	0	0
130	0,65	0,6	0,6	0,6	0,8	0,85	0,85	0,75	0,7	0	0,15
140	0,6	0,6	0,65	0,65	0,75	0,8	0,75	0,65	0,5	0	0,1
150	0,55	0,65	0,6	0,65	0,65	0,65	0,6	0,65	0,5	0	0

Дані таблиці 4.3 показують, що *НМ_v2*, по-перше, майже повністю виправила помилку під час визначення моделі дрону (комірки позначені *), а, по-друге, значно покращила ймовірність ефективного виявлення на великій відстані залежно від рівня шуму оточуючого середовища. Представимо результати графічно (рис. 4.13 – 4.23).

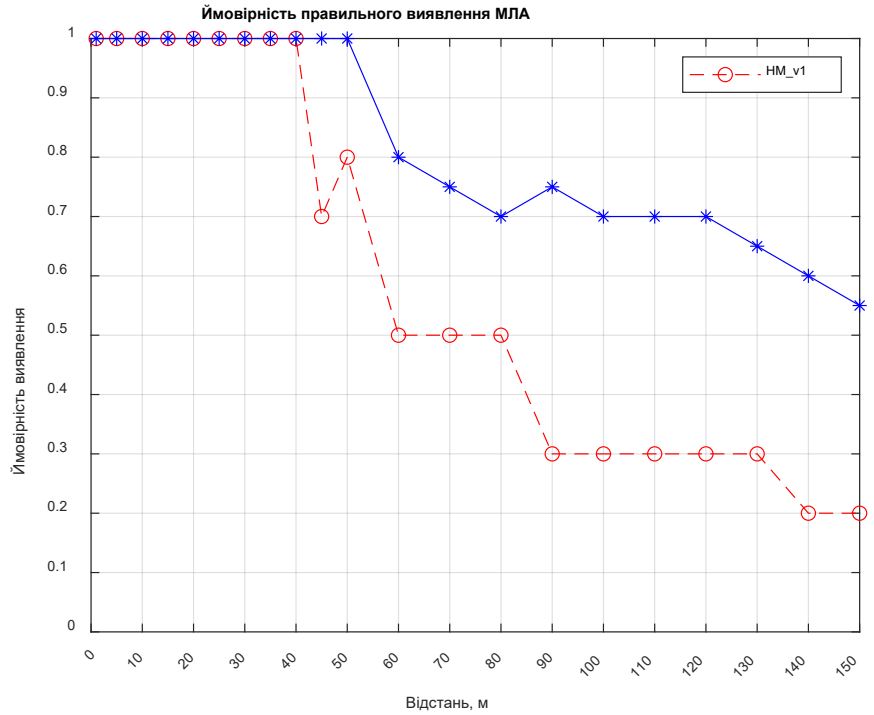


Рисунок 4.13 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 40 дБ

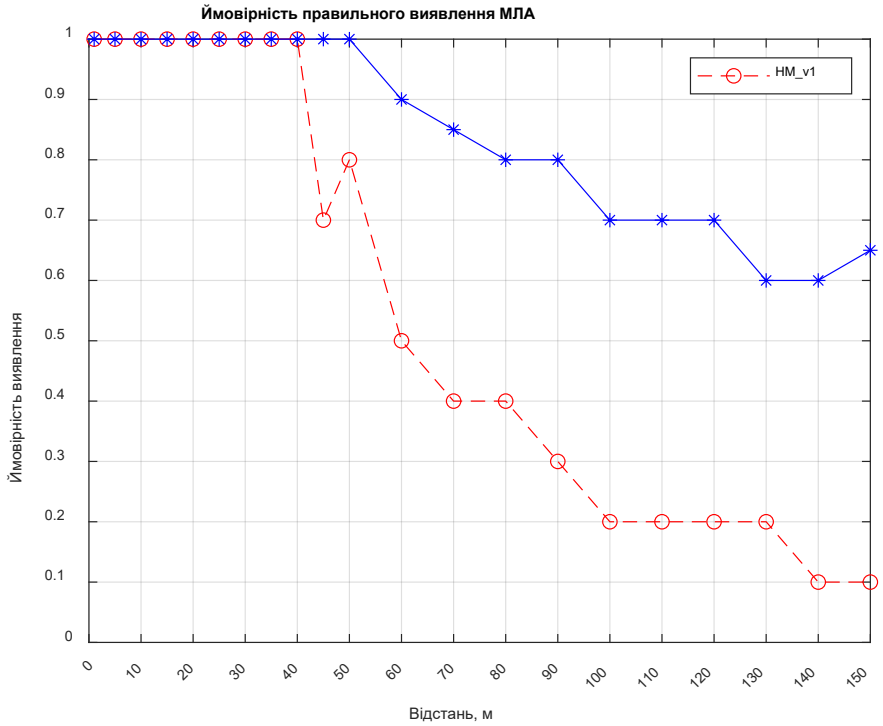


Рисунок 4.14 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 37 дБ

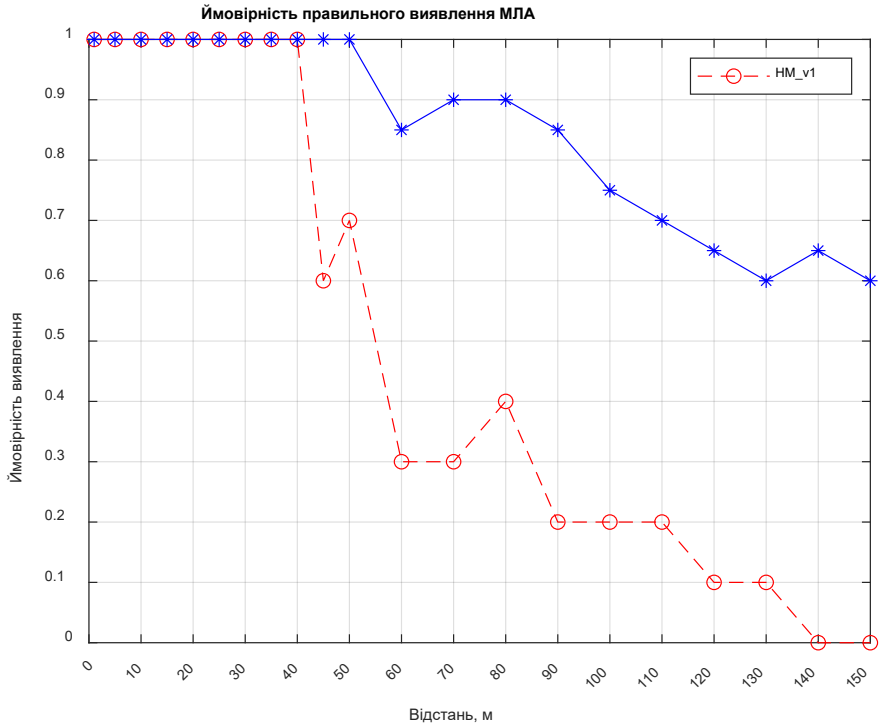


Рисунок 4.15 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 35 дБ

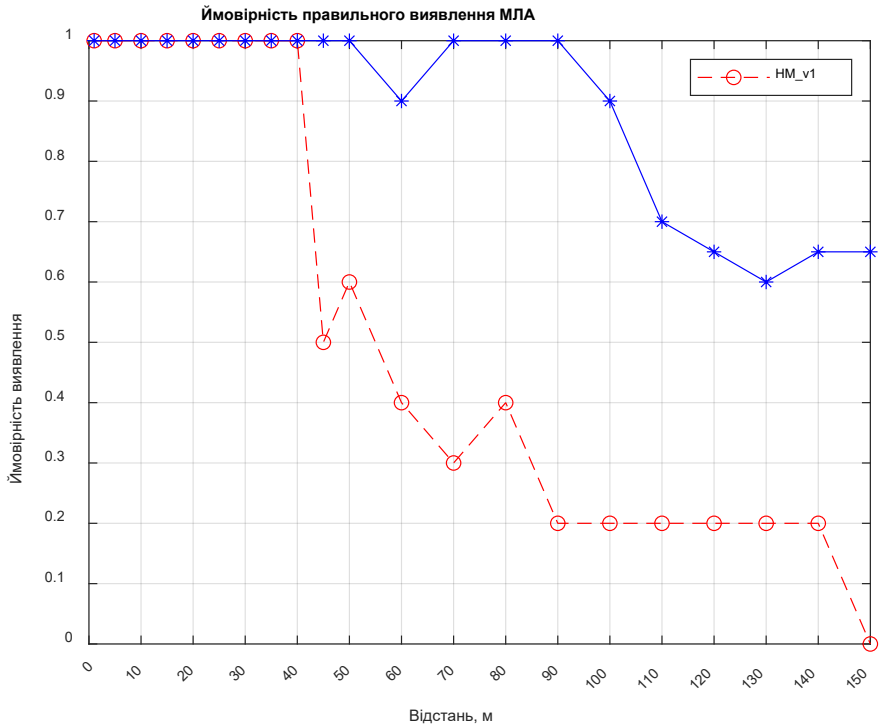


Рисунок 4.16 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 33 дБ

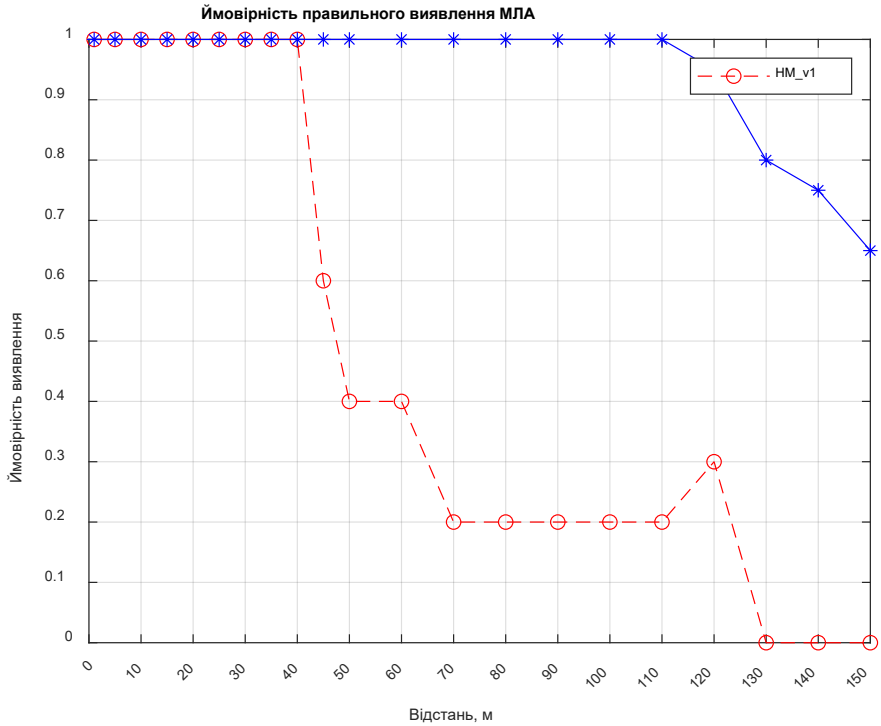


Рисунок 4.17 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 30 дБ

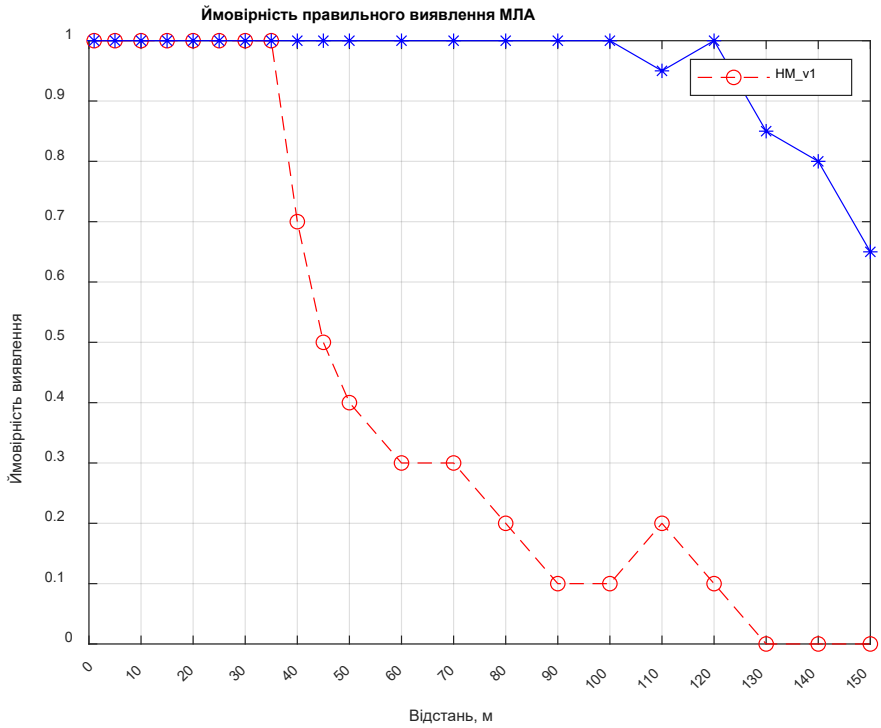


Рисунок 4.18 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 27 дБ

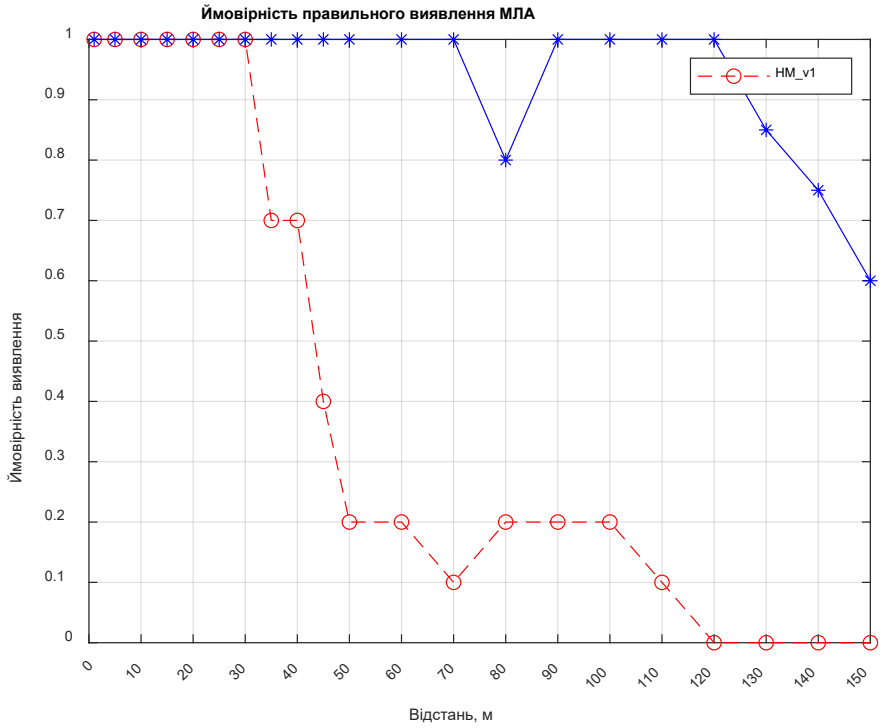


Рисунок 4.19 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 24 дБ

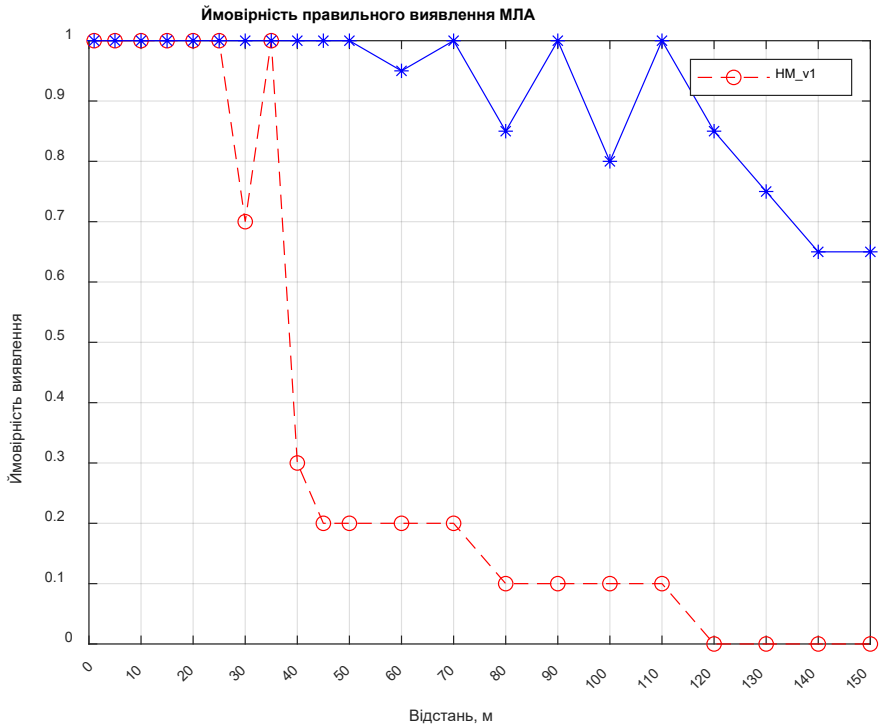


Рисунок 4.20 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 21 дБ

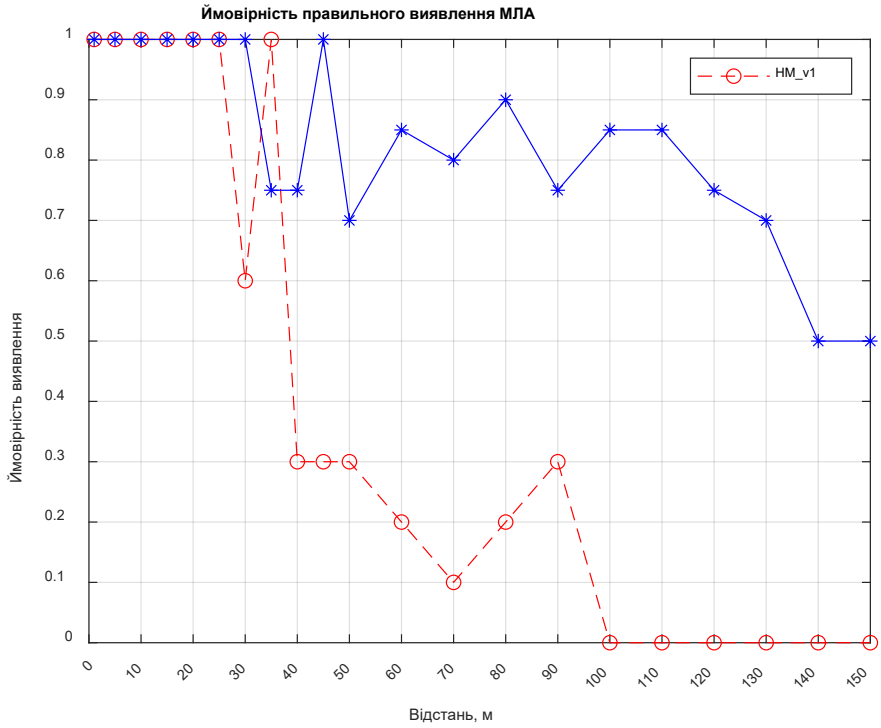


Рисунок 4.21 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 18 дБ

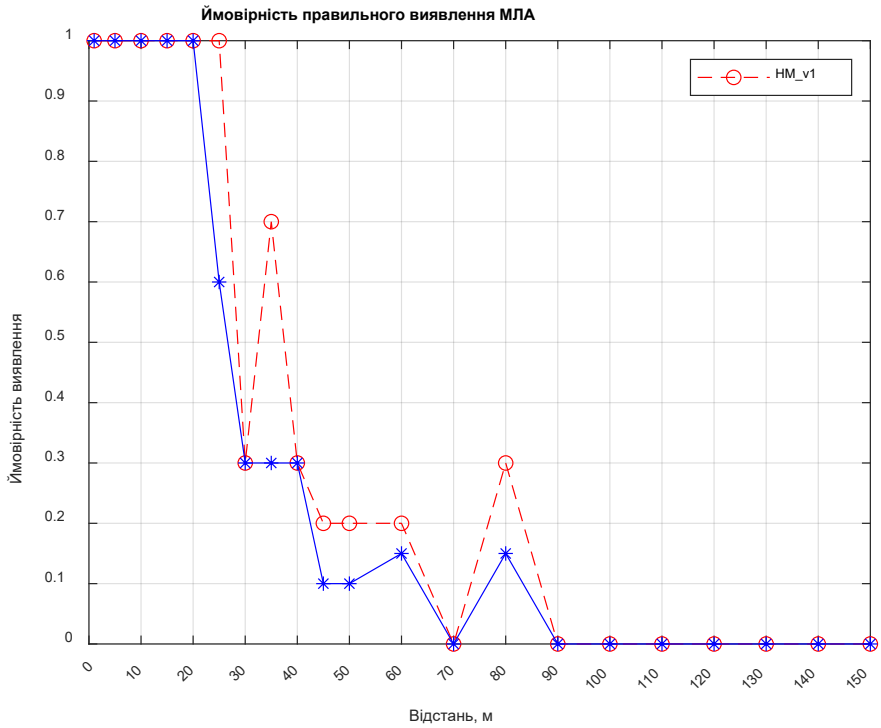


Рисунок 4.22 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 15 дБ

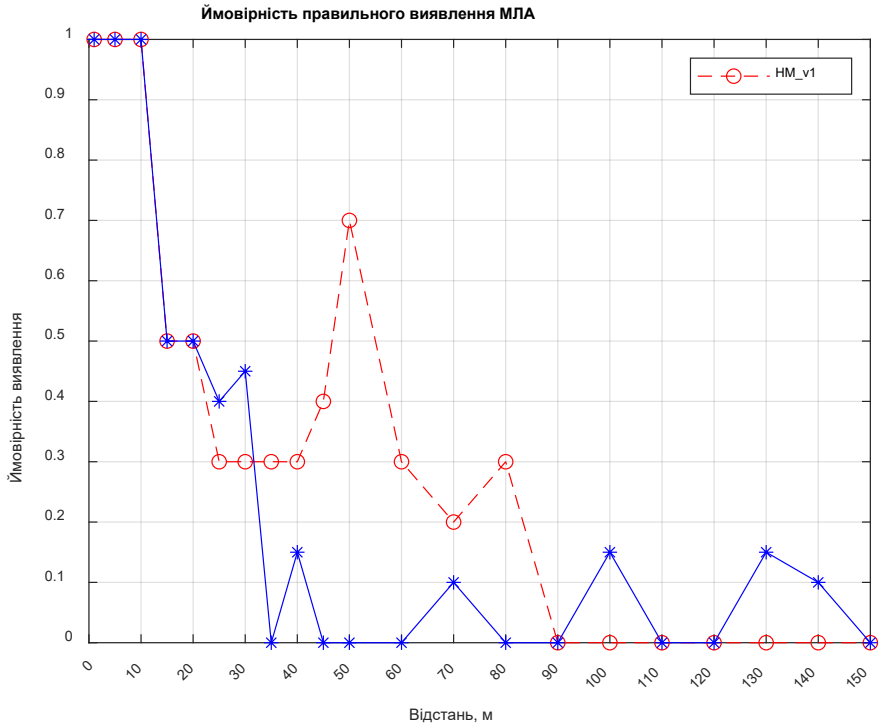


Рисунок 4.23 — Ймовірність виявлення МЛА алгоритмом за умови С/Ш 12 дБ

4.2. Рекомендації для підвищення ефективності роботи системи виявлення МЛА

Крім збільшення бази аудіоданих, для покращення роботи системи виявлення МЛА можна працювати ще в декількох напрямах:

По-перше, це збільшення глибини архітектури розробленої згорткової нейронної мережі. У зв’язку зі збільшенням бази навчальних аудіосигналів, зростає кількість пар шарів згортки та максимального об'єднання в архітектурі ЗНМ для ефективного вивчення ознак вищої складності та абстракції, а також зменшення просторових розмірів. За кожної зміни в структурі моделі, мережу необхідно заново перенавчати.

По-друге, покращення приймального обладнання, збільшує якість та максимальну відстань зареєстрованих звукових хвиль від безпілотних літальних апаратів. У якості приймальних пристроїв таких систем краще використовувати професійні високочутливі мікрофони, з великим динамічним діапазоном та захистом від шумів вітру. Такі мікрофони бажано об’єднувати в групи для перекриття більшої ефективної площі, утворюючі мікрофонні масиви.

Додатковим обладнанням до електроакустичного тракту детектора є узгоджуючи акустичні фільтри для зменшення рівня завад.

Залежно від обраної моделі мікрофону, що відрізняється за характеристиками від розглянутої в цій роботі, необхідно перероблювати приймальний тракт, наприклад, використати передпідсилювачі, смугові фільтри або обрати АЦП із більш високою розрядністю, перерахувавши його основні параметри. Усі необхідні рекомендації надано в розділі 2.

По-третє, застосування даних з інших типів сенсорів, таких як камери візуального спостереження або радіолокаційні станції, поряд з акустичним аналізом, забезпечує мультимодальний підхід до виявлення безпілотників. Поєднуючи різні модальності сенсорів, алгоритм може використовувати додаткову інформацію і підвищити загальну точність виявлення, особливо в сценаріях, де одного лише звуку може бути недостатньо.

По-четверте, вирішальне значення мають безперервне вдосконалення та оптимізація алгоритму. Регулярні оновлення, які ґрунтуються на спостережені, аналізі ефективності роботи та поточних дослідженнях, гарантують, що алгоритм залишатиметься адаптивним до нових технологій МЛА, їхніх потенційних ворожих атак і маневрів ухилення від виявлення.

По-п’яте, можна дослідити методи машинного навчання, які дають змогу алгоритму адаптуватися і навчатися в режимі реального часу. Підходи онлайн-навчання, коли алгоритм оновлює свої можливості виявлення в міру того, як він стикається з новими даними, можуть підвищити його гнучкість і адаптивність до нових загроз, пов'язаних із безпілотниками. Крім того, використання можливостей хмарних обчислень і розподілених систем може уможливити швидшу обробку й аналіз великих наборів аудіоданих із дронів, що полегшить виявлення і класифікацію в реальному часі.

Не менш важливим способом вдосконалення алгоритму є співпраця та обмін даними між дослідниками та зацікавленими сторонами в цій галузі. Створюючи відкриті набори даних, дослідники отримують доступ до ширшого

спектру аудіозаписів дронів, що дасть їм нагоду розробляти більш надійні й точні алгоритми, використовуючи колективні знання та ресурси.

Висновки до розділу 4

1. Було проведено тестування роботи готової моделі згорткової нейронної мережі для виявлення та класифікації малих безпілотних літальних апаратів.

2. Ефективна дистанція виявлення алгоритму дорівнює 200 метрів, що за методом експертних оцінок, значно перевершує можливості людського слуху з максимальною відстанню до 120 метрів.

3. Надані рекомендації для подальшого розвитку та підвищення ефективності роботи системи розпізнавання сигналів від малих безпілотних літальних апаратів.

4. Основні результати тестування виявлення квадрокоптера моделі *«DJI Mavic 3»* розробленою системою, описані у розділі 4, зазначено автором у науковій статті [50].

ЗАГАЛЬНІ ВИСНОВКИ

У дисертаційній роботі вирішено важливе та актуальне науково-технічне завдання — розроблення пристрою з одним вхідним сенсором-мікрофоном для виявлення малих безпілотних літальних апаратів на основі аналізу акустичних сигнатур. Розроблена модель системи підвищує ефективність виявлення дронів за рахунок використання нейронної мережі для обробки акустичних сигналів.

У результаті проведення досліджень отримано такі наукові та практичні результати:

1. Під час проведення аналізу основних методів виявлення та локалізації малих безпілотних літальних апаратів було визначено, що серед усіх методів, найбільш перспективним є акустичний. Теоретична ефективна дальність виявлення цілі таким видом детектора, на прикладі моделі дрону «*Mavic 3*», складає 78 м.

2. За час дослідження була створена база аудіозаписів акустичних сигнатур малих безпілотних літальних апаратів моделей «*DJI Mavic 3*», «*DJI Mavic 2 Pro*», «*FPV*», *«FeiLun FX137»* та «*LH X43-W*».

3. Розроблено математичну модель для ідентифікації типу малих безпілотних літальних апаратів, використовуючи базу аудіозаписів акустичних сигнатур.

4. Розроблено метод цифрової обробки акустичних сигналів із використанням згорткової нейронної мережі глибокого навчання, що складається з 10 пар шарів згортки, ReLU, пакетної нормалізації і максимального об'єднання, а також одного шару згладжування, відсікання, повнозв'язного, Softmax та фінального шару класифікації. Для ефективного навчання моделі застосовано оптимізатор Adam.

5. На основі запропонованого методу, розроблено програмне забезпечення для швидкої та ефективної обробки акустичних сигналів на предмет наявності в них сигнатур малих безпілотних літальних апаратів.

6. Створено акустичний детектор малих безпілотних літальних апаратів. Розроблений у роботі пристрій складається з електротного мікрофону моделі *CMA-4544PF-W*, передпідсилювача та підсилювача із системою АРП на базі мікросхеми *MAX9814*, 14-бітного АЦП із частотою дискретизації не менше 40 кГц та зовнішньою звуковою картою *CM6206* із лінійним входом, що під'єднана до ПК, де і відбувається основна обробка акустичних сигналів за допомогою розробленого спеціального ПЗ.

7. Під час перевірки роботи апаратного забезпечення було визначено, що ефективна дистанція виявлення МЛА пристрою із розробленим програмним забезпеченням, на прикладі тієї ж моделі дрону «*DJI Mavic 3*», складає 200 м. Цей показник майже в 3 рази перевищує теоретичні розрахунки (78 м) та у 2 рази — можливості людського слуху (120 м).

8. Надано універсальні рекомендації для подальшого розвитку та підвищення ефективності роботи системи виявлення та локалізації малих безпілотних літальних апаратів.

Таким чином, можна стверджувати, що поставлені завдання виконано в повному, отже основна мета дисертаційної роботи досягнута.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Szabolcs Prisznyak, «Drones and Jails», *Buletin Stintific*, vol. 23, iss. 1, pp. 43 – 52, July. 2018. doi: 10.2478/bsaft-2018-0006

[2] K. Dalamagkidis, “Classification of UAVs” in *Handbook of Unmanned Aerial Vehicles*, K. P. Valavanis and G. J. Vachtsevanos, Denver, CO, USA:Springer Dordrecht, 2015, pp. 83-91.

[3] В. М. Карташов, В. Н. Олейников, С. А. Шейко, С. И. Бабкин, И. В. Корытцев, и О. В. Зубков, «Особенности обнаружения и распознавания малых беспилотных летательных аппаратов», *Радиотехника : Всеукр. межвед. науч.-техн. сб,* №195, с. 235 – 243, 2018.

[4] Ю. Г. Даник та М. В. Бугайов, «Аналіз ефективності виявлення тактичних безпілотних літальних апаратів пасивними та активними засобами спостереження», *зб. наук. пр. "Проблеми створення, випробування, застосування та експлуатації складних інформаційних систем"*, №10, с. 5 – 20, 2015.

[5] А. Я. Торопчин та ін., *Довідник з протиповітряної оборони.* Київ, Україна: МО України, 2003.

[6] Я. І. Лепіх та ін., *Оптико-електронні системи ближньої локації: монографія.* Одеса, Україна: Одес. нац. ун-т ім. І. І. Мечникова, 2019.

[7] О.І. Сухаревський, В.О. Василець, Я.О. Белевщук та К.І. Ткачук, «Розрахунок радіолокаційних характеристик моделі безпілотного літального апарату RQ-1 “PREDATOR”», *Системи обробки інформації.* Вип. 5 (103), с. 68-72, 2012.

[8] DN-9770D - камера відеоспостереження вулична. [Електронний ресурс]. Доступно: http://www.rikas-varta.com.ua/index.php?inx=prd.cam.dn&prt=DN-9770D#. Дата звернення: Бер. 10, 2021.

[9] R. Austin, “Design for Stealth”, in *Unmanned aircraft systems: UAVS design, development and deployment*. Chichester, UK: John Wiley and Sons Ltd, 2010, pp. 113–127.

[10] C. G. Ribbing, «Controlling thermal radiation from surfaces», in *Optical Thin Films and Coatings*, A. Piegari and F. Flory, Sawston, UK: Woodhead Publishing Limited, 2013, pp. 811-835. doi:10.1533/9780857097316.2.357.

[11] *Таблиця спектральних коефіцієнтів випромінювання,* 2016. [Електронний ресурс]. Доступно: https://www.bergeng.com/mm5/downloads/fluke/ Emissivity-Values-of-Common-Materials-Chart.pdf. Дата звернення: Січ. 19, 2020.

[12] Д. В. Шимонець, О. Л. Лугина, М. Ю. Харьков, І. В. Нечитайло, та Г.В.Худов, «Пасивні радіолокаційні системи виявлення малорозмірних повітряних об'єктів з використанням сигналів телекомунікаційних систем», *Системи управління, навігації та зв'язку. Збірник наукових праць,* №4 (44), с. 138-142, 2017.

[13] Д. І. Вельміскін, *Теоретичні основи радіолокації: конспект лекцій.* Дніпропетровськ, Україна: "Економіка", 2005.

[14] J. J. Beel, *«Anti-UAV Defense For Ground Forces and Hypervelocity Rocket Lethality Models»*, M.S. thesis, Naval Postgraduate School, Monterey, California, USA, 1992. pp. 36–46.

[15] К. С. Васюта, О. В. Тесленко, В. М. Купрій, та О. А. Малишев, «*Основи побудови радіолокаційних засобів розвідки повітряного простору: конспект лекцій*. Харків, Україна: ХУПС, 2013.

[16] А. В. Коротун, Н. А. Смирнова та Я. В. Карандась, *Методичні вказівки до практичних занять та самостійної роботи з дисципліни «Мікро- і наносенсори» для студентів спеціальності 153 «Мікро- та наносистемна техніка», освітня програма «Мікро- та наноелектронні прилади та пристрої» денної й заочної форм навчання*. Запоріжжя, Україна: НУ «Запорізька політехніка», 2019.

[17] В. Ф. Самохин, С. П. Остроухов и П. А. Мошков, «Экспериментальное исследование источников шумности беспилотного летательного аппарата с винто-кольцевым движителем в толкающей компоновке», *Электронный журнал «Труды МАИ»*, № 70, с.1 – 24, 2012.

[18] С.О. Козерук та О.В. Коржик, «Виявлення малих літальних апаратів за акустичним випромінюванням», *Вісник НТУУ "КПІ". Серія Радіотехніка, Радіоапаратобудування,* №76, с.15–20, 2019. doi:10.20535/RADAP.2019.76.15-20

[19] Г. В. Пєвцов, А. Я. Яцуценко, М. Ф. Пічугін, Д. В. Карлов, Ю. В. Трофіменко, та М. В. Борцова, «Розвиток теорії виявлення радіосигналів. Основи енергетичного виявлення», *Озброєння та військова техніка.* №3, с. 24-36, 2015.

[20] С. О. Сокольський та А. В. Мовчанюк, «Огляд методів виявлення та локалізації малих безпілотних літальних апаратів», *Вісник НТУУ "КПІ". Серія Радіотехніка. Радіоапаратобудування,* №87, с. 46-55, Грудень. 2021. doi:10.20535/RADAP.2021.87.46-55.

[21] Datasheet *CMA-4544PF-W.* [Електронний ресурс]. Доступно: https://octopart.com/datasheet/cma-4544pf-w-cui+devices-106233423. Дата звернення: Бер. 10, 2023.

[22] Р.Б. Небожук, «Метод виділення огинаючої складової мовних сигналів для ефективного їхнього кодування», кваліфікаційна робота магістра, Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, Україна, 2019.

[23] О.П. Гребінь, Н.Ф. Левенець та В.Б. Швайченко, *Прикладна акустика – 1. Електроакустика: Навчальний посібник: навч. посіб. для студ. спеціальності 171 «Електроніка», спеціалізації «Електронні та інформаційні технології кінематографії та аудіовізуальних систем»*. Київ, Україна: КПІ ім. Ігоря Сікорського, 2018.

[24] *Reclamination Managing Water in the West. Resource management plan Navajo reservoir area Colorado and New Mexico. Final environmental assessment and finding of no significant impact*. Colorado, USA: Navajo Reservoir RMP/FEA, June. 2008. pp.295-296.

[25] М. Є. Сухов, С. Д. Бать, В. В. Колосов та О. Г. Чупаков*, Схемотехніка високоякісного звуковідтворення*. Київ, Україна: Техніка, 1992.

[26] P. Skritek, *Handbuch der Audio-Schaltungstechnik: Berechnungsmethoden, Messverfahren, Schaltungsentwicklung, digitale Tonverarbeitung*. München, Deutschland: Franzis, 1988.

[27] P. Horowitz and W. Hill, *The Art of Electronics*, Third Edition. , New York, USA: Cambridge University Press, 2015.

[28] І.М. Миценко и Д.Д. Халамейда, «Импульсный логарифмический усилитель с автоматической регулировкой усиления в диапазоне частот 1 МГц…8 ГГц», *Радіофізика та електроніка*, т. 23, № 1, с. 4-9, 2018. doi: 10.15407/rej2018.01.004

[29] В. В. Гриненко, *Пристрої аналогової електроніки : конспект лекцій*. Суми, Україна : Сумський державний університет, 2015.

[30] *Datasheet MAX9814.* [Електронний ресурс]. Доступно: https://datasheets.maximintegrated.com/en/ds/MAX9814.pdf . Дата звернення: Бер. 19, 2023.

[31] *Datasheet CM6206.* [Електронний ресурс]. Доступно: https://datasheetspdf.com/pdf-file/718705/CEDIAELECTRONICS/CM6206 Дата звернення: Бер. 19, 2023.

[32] С.О. Козерук та О.В. Коржик, «Виявлення, локалізація та ідентифікація малих літальних апаратів за акустичним випромінюванням», *Вісник НТУУ "КПІ". Серія Радіотехника. Радіоапаратобудування*, № 89, с. 29 – 38, Вересень. 2022. DOI: 10.20535/RADAP.2022.89.29-38

[33] В. М. Карташов та ін. «Информационные характеристики звукового излучения малых беспилотных летательных аппаратов», *Радиотехника: Всеукр. межвед. науч.-техн. сб*, №191, с. 181-187, 2017.

[34] Мінрегіон України. 1 сесія. (2013, груд. 27). *Наказ № 630, ДБН В.1.1-31:2013 Захист територій, будинків і споруд від шуму*. [Електронний ресурс]. Доступно: https://dbn.co.ua/load/0-0-0-1814-20. Дата звернення: Бер. 23, 2023.

[35] С. О. Сокольський та А. В. Мовчанюк, «Електроакустичний тракт детектора для виявлення малих безпілотних літальних апаратів», *Вісник*

Вінницького політехнічного інституту, № 2, с. 135-144, Травень. 2023. doi: 10.31649/1997-9266-2023-167-2-135-144.

[36] Г. С. Ігнатенко та А. Г. Ламчановський, «Класифікація аудіосигналів за допомогою нейронних мереж», *Міжнародний журнал «Молодий вчений»*, №48(286), с. 23-25, 2019.

[37] J. Singh, «An introduction to audio processing and machine learning using Python», 2019. [Online]. Avaliable: https://opensource.com/article/19/9/audio-processing-machine-learning-python. Accessed on: May 19, 2019.

[38] N. Pratheeksha, *«The dummy's guide to MFCC»,* 2018. [Online]. Avaliable: https://medium.com/prathena/the-dummys-guide-to-mfcc-aceab2450fd. Accessed on: May 23, 2021.

[39] R. Callan, *The Essence of Neural Networks.* London, England: Pearson Education, 2010.

[40] С. Петров, «Сверточная нейронная сеть для распознавания символов номерного знака автомобиля», *Международный Университет природы, общества и человека «Дубна»*, №3, с. 56, 2013.

[41] G.V. Cybenko, «Approximation by superpositions of a sigmoidal function», *Mathematics of Control, Signals, and Systems. Springer International*, №2, pp. 303 – 314, 1989.

[42] H. Wu, «Global stability analysis of a general class of discontinuous neural networks with linear growth activation functions», *Information Sciences*, vol. 179, iss. 19, pp. 3432–344, 2009. doi:10.1016/j.ins.2009.06.006.

[43] D. Sussillo and L. F. Abbott, «Random Walk Initialization for Training Very Deep Feedforward Networks», in *3rd International Conference on Learning Representations*, San Diego, CA, USA, 2015, pp. 1 – 10. doi:10.48550/arXiv. 1412.6558.

[44] V. Nair and G. E. Hinton, «Rectified Linear Units Improve Restricted Boltzmann Machines», in *27th International Conference on International Conference on Machine Learning*, Haifa, Israel, 2010, pp. 807–814. doi:10.5555 /3104322.3104425.

[45] A. L. Maas, A. Y. Hannun and A. Y. Ng, “Rectifier nonlinearities improve neural network acoustic models”, in *30th International Conference on Machine Learning*, vol. 28, №3, Atlanta, USA, 2013, pp. 1 – 6.

[46] D. A. Clevert, T. Unterthiner and S. Hochreiter, “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)” *in 4th International Conference on Learning Representations*, San Juan, Puerto Rico, 2016, pp.1–14. doi:10.48550/arXiv.1511.07289.

[47] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville and Y. Bengio, “Maxout Networks” in *30th International Conference on Machine Learning*, vol. 28, Atlanta, USA, 2013, pp. 1 – 9. doi:10.48550/arXiv.1302.4389.

[48] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2012.

[49] D. P. Kingma and J. Ba, «Adam: A method for stochastic optimization», in *3rd International Conference for Learning Representations*, San Diego, USA, 2015, pp. 1–15. doi:10.48550/arXiv.1412.6980.

[50] С. О. Сокольський та А.В. Мовчанюк, «Алгоритм оброблення аудіосигналів із використанням методу машинного навчання», *Вісник НТУУ "КПІ". Серія Радіотехніка. Радіоапаратобудування,* №93, с. 39-51, Вересень. 2023. doi:10.20535/RADAP.2023.93.39-51

ДОДАТОК А

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ ТА ВІДОМОСТІ ПРО АПРОБАЦІЮ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЇ

Наукові праці, в яких опубліковані основні наукові результати дисертації

1. С. О. Сокольський та А. В. Мовчанюк, «Огляд методів виявлення та локалізації малих безпілотних літальних апаратів», *Вісник НТУУ "КПІ". Серія Радіотехніка. Радіоапаратобудування,* №87, с. 46-55, Грудень. 2021. doi:10.20535/RADAP.2021.87.46-55 (фахове видання категорії А).
2. С. О. Сокольський та А. В. Мовчанюк, «Електроакустичний тракт детектора для виявлення малих безпілотних літальних апаратів», *Вісник Вінницького політехнічного інституту*, № 2, с. 135-144, Травень. 2023. doi: 10.31649/1997-9266-2023-167-2-135-144 (фахове видання категорії Б).
3. С. О. Сокольський та А.В. Мовчанюк, «Алгоритм оброблення аудіосигналів із використанням методу машинного навчання», *Вісник НТУУ "КПІ". Серія Радіотехніка. Радіоапаратобудування,* №93, с. 39-51, Вересень. 2023. doi:10.20535/RADAP.2023.93.39-51 (фахове видання категорії А).

Наукові праці, які засвідчують апробацію матеріалів дисертації

1. С.О. Григораш (Сокольський) та А.В. Мовчанюк, «Нові виміри застосування дронів», на III *Міжнародній науково-практичній конференції "Філософія і наукова-технічна творчість у хронотопі технічного університету"*, Київ, Україна, 2020, с. 92-97 (матеріали конференції)
2. С.О. Сокольський та А.В. Мовчанюк, «Огляд методів виявлення і локалізації МЛА», на *X Міжнародній науково-технічній конференції «РТПСАС-2021*», Київ, Україна, 2021, с. 48-50 (матеріали конференції)
3. С.О. Сокольський та А.В. Мовчанюк, «Електроакустичний тракт детектора МЛА», на *XI Міжнародній науково-технічній конференції «РТПСАС-2022*», Київ, Україна, 2022, с. 88-90 (матеріали конференції)

ДОДАТОК Б
АКТИ ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ ДИСЕРТАЦІЙНОЇ РОБОТИ

ДОДАТОК В
КОД ПРОГРАМНОГО АЛГОРИТМУ ВИЯВЛЕННЯ ТА КЛАСИФІКАЦІЇ СИГНАЛІВ МАЛИХ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ

Основний код програми у файлі ***main_project.mxl***:

В.1. Завантаження Аудіо даних

Вказати шлях до даних проекту.

```
projectFolder = 'C:\Users\sergi\Documents\MATLAB\Examples\R2022a\deep-learning-sound-recognition_PhDproj';
dataset = fullfile(projectFolder,'AudioDataStore','dron_signals'); %папка з аудіо файлами
```

В.2. Створення сховища тренувальних даних (Training Datastore)

Створимо *audioDatastore*, який вказує на набір даних для навчання *(training data set)*.

AudioDatastore для керування колекцією шумів МЛА створимо так, щоб кожна підпапка колекції названа на честь відповідної моделі МЛА, яку вона представляє. Встановимо властивість *'IncludeSubFolders'* у значення *true*, щоб доручити сховищу аудіо даних використовувати підтеки (підпапки), і встановимо властивість *'LabelSource'* у значення *'foldernames'*, щоб створити мітки даних на основі назв підпапки теки.

```
ads = audioDatastore(fullfile(dataset,"train"), ...
   IncludeSubfolders=true, ...
   FileExtensions=".wav", ...
   LabelSource="foldernames")
```

В.3. Вибір моделей МЛА для виявлення

Задамо назви моделей дронів, які система має розпізнавати.

```
commands = categorical(["mavic3", "Mavic2_Pro","feilun_fx137"]);
```

Позначимо усі інші категорії, які не є у масиві commands, як невідомі (unknown). Мережа використовує цю групу, щоб дізнатися різницю між командами і всіма іншими файлами.

isCommand = ismember(ads.Labels,commands); %перевіряє елементи масиву ads на наявність хоча б одного зі значень масиву commands
isUnknown = ~isCommand; %інші дані помічає як невідомі

Щоб зменшити дисбаланс класів між відомими і невідомими категоріями та пришвидшити обробку, можна включити до навчального набору лише частину невідомих слів.

includeFraction = 1;%відсоток даних для перевірки (зараз 100%)

Вектор *mask* є двійковою маскою, яку можна використовувати для випадкового вибору підмножини даних з ads з ймовірністю includeFraction. Якщо includeFraction встановлено на 0.5, то приблизно половина елементів у результуючому логічному векторі буде істинною, і відповідні елементи в adsTrain будуть вибрані з ймовірністю 50%.

mask = rand(numel(ads.Labels),1) < includeFraction; %повертає матрицю (розмір: кількість елементів у масиві ads.Labels x 1) рандомних значень, в якій елементи, менші за порогове значення includeFraction, отримують значення true, а решта - false.

isUnknown & mask — виконує логічне перемноження масивів isUnknown та mask і повертає масив, що містить елементи, які дорівнюють або логічній 1 (істина), або логічному 0 (брехня). Якщо елементи при множенні ненульові, то у результуючому масиві елемент буде дорівнювати логічній 1 (true), в іншому випадку елемент масиву дорівнює 0.

isUnknown = isUnknown & mask; % логічне І масивів isUnknown та mask
ads.Labels(isUnknown) = categorical("unknown"); %додаємо нову категорію "unknown" з даними із ads.Labels(isUnknown)
adsTrain = subset(ads,isCommand|isUnknown); %повертає підмножину, що містить тільки команди та невідомі слова
countEachLabel(adsTrain) %кількість прикладів, що належать для кожної категорії

В.4. Створення валідаційної бази даних (Validation Datastore)

Створимо *audioDatastore*, який вказує на набір даних валідації.

```
ads = audioDatastore(fullfile(dataset,"validation"), ...
        IncludeSubfolders=true, ...
        FileExtensions=".wav", ...
        LabelSource="foldernames")
isCommand = ismember(ads.Labels,commands);
isUnknown = ~isCommand;
includeFraction = 1;
mask = rand(numel(ads.Labels),1) < includeFraction;
isUnknown = isUnknown & mask;
ads.Labels(isUnknown) = categorical("unknown");
adsValidation = subset(ads,isCommand|isUnknown);
countEachLabel(adsValidation)
```

В.5. Обчислення слухових спектрограм

Щоб підготувати дані для ефективного навчання згорткової нейронної мережі, перетворимо вхідні аудіофайли на слухові спектрограми.

Напишемо для цього спеціальну функцію ***ExtractAuditoryFeatures,*** яка приймає на вхід 2 аргументи: вхідний сигнал (x) та частоту дискретизації (Fs).

```
Fs = 48e3; % Частота дискретизації
segmentDuration = 1; % тривалість кожного мовного кліпу (у секундах)
frameDuration = 0.025; % тривалість кожного кадру для розрахунку спектру
hopDuration = 0.010; %часовий крок між спектрами
numBands = 30; %кількість фільтрів у слуховій спектрограмі.

segmentSamples = round(segmentDuration*Fs);
frameSamples = round(frameDuration*Fs);
hopSamples = round(hopDuration*Fs);
overlapSamples = frameSamples - hopSamples;
```

Довжина ШПФ повинна бути принаймні такою ж, як вікно аналізу, щоб запобігти витоку спектра і досягти хорошої частотної роздільної здатності. У випадку оптимальною довжиною ШПФ зазвичай є степінь двійки, яка більша або дорівнює довжині вікна Хеммінга. Наприклад, якщо вікно Хеммінга має довжину 256 відліків, то відповідною довжиною ШПФ буде 512 або 1024 відліки.

```
% FFTLength = frameSamples;

FFTLength = 1200;

aFE = audioFeatureExtractor( ...
   SampleRate=Fs, ...
   FFTLength=FFTLength, ...
   Window=hamming(frameSamples,'periodic'), ... % Hamming window with length (frameSamples) and "periodic" in type
   OverlapLength=overlapSamples, ...  % Довжина (overlapSamples) перекриття сусідніх вікон аналізу
```

melSpectrum=true); % Мел спектр

% Для встановлення параметрів видобування мел-спектра використовуйте setExtractorParameters:

setExtractorParameters(aFE,"melSpectrum",NumBands=numBands,WindowNormalization=false);

Прочитаємо файл з набору даних.

Для навчання згорткової нейронної мережі потрібно, щоб вхідні дані були однакового розміру. Деякі файли у наборі даних мають довжину менше 1 секунди. Застосуємо нульове заповнення до передньої та задньої частин аудіосигналу, щоб він мав довжину segmentSamples.

x = read(adsTrain);
numSamples = size(x,1); %повертає 1-ий елемент вектора розміру масива adsTrain (Files)
numToPadFront = floor((segmentSamples - numSamples)/2); % Округлення у менший бік
numToPadBack = ceil((segmentSamples - numSamples)/2); % Округлення у більший бік
xPadded = [zeros(numToPadFront,1,"like",x);x;zeros(numToPadBack,1,"like",x)]; %доповнення вхідного сигналу x нулями на початку і в кінці, щоб збільшити його довжину на задану величину. Результуючий доповнений сигнал зберігається у змінній xPadded.

Щоб витягти звукові характеристики, викличемо команду extract. На виході ми отримуємо Мел-спектр з часом по рядках.

features = extract(aFE,xPadded); % повертає всі підрядки в рядку aFE, які відповідають шаблону, заданому в xPadded

[numHops,numFeatures] = size(features) % присвоює змінній numHops кількість рядків матриці ознак, а змінній numFeatures - кількість стовпців.

- numHops представляє кількість "стрибків" або часових кроків у матриці ознак. Кожен рядок матриці містить набір ознак, витягнутих з невеликого вікна аудіо.
- numFeatures - кількість ознак, вилучених для кожного часового кроку. Це можуть бути спектральні ознаки, такі як MFCC, спектральний потік або кольоровість, а також будь-які інші відповідні ознаки, що використовуються для аналізу звуку.

Наступний фрагмент коду перевіряє, чи доступний *Parallel Computing Toolbox* і чи не встановлено прапорець *reduceDataset*. Якщо обидві умови істинні, створюється паралельний пул робітників і визначається кількість розділів, на які можна розділити дані. Якщо будь-яка з умов є хибною, кількість розділів встановлюється рівною 1, що означає, що дані будуть оброблятися послідовно.

if ~isempty(ver("parallel")) && ~speedupExample % ver('parallel') перевіряє, чи встановлено Parallel Computing Toolbox

pool = gcp; %створення паралельного пулу робітників

numPar = numpartitions(adsTrain,pool); % кількість розділів, на які можна розбити дані за допомогою функції numpartitions

else

% ОБИРАЄ ЦЕ, оскільки Parallel Computing Toolbox НЕ інстальований

numPar = 1; %кількість розділів встановлюється рівною 1 => дані обробляються послідовно

end

Для кожного розділу зчитуємо зі сховища даних, обнуляємо сигнал, а потім виділяємо ознаки

parfor ii = 1:numPar % паралельний цикл for, який перебирає кількість розділів, на які були розбиті дані (numPar).

% Кожна ітерація циклу обробляє підмножину навчальних даних паралельно на окремому робочому елементі.

subds = partition(adsTrain,numPar,ii); %розбиває сховище даних adsTrain на numPar рівних за розміром підмножин і вибирає і-ту підмножину для обробки.

XTrain = zeros(numHops,numBands,1,numel(subds.Files)); %Створюється заповнений нулями багатовимірний масив XTrain для зберігання вилучених ознак.

for idx = 1:numel(subds.Files) %цикл перебирає кожен файл у поточній підмножині навчальних даних.

x = read(subds); %Зчитує аудіосигнал з поточного файлу у змінну x

xPadded = [zeros(floor((segmentSamples-size(x,1))/2),1);x;zeros(ceil((segmentSamples-size(x,1))/2),1)]; %доповнюємо аудіосигнал нулями на початку і в кінці

XTrain(:,:,:,idx) = extract(aFE,xPadded); % Витягує ознаки з доданого аудіосигналу за допомогою екстрактора

end

XTrainC{ii} = XTrain; %Зберігає вилучені ознаки з поточної підмножини навчальних даних у масиві комірок XTrainC, індексованому поточною ітерацією циклу

end

Перетворимо вихідні дані в 4-вимірний масив зі слуховими спектрограмами вздовж четвертого виміру.

XTrain = cat(4,XTrainC{:}); %Конкатенація (об'єднання) вилучених ознак з усіх підмножин навчальних даних вздовж четвертого виміру, в результаті чого утворюється 4D масив XTrain з розмірами numHops x numBands x 1 x numSpec.
[numHops,numBands,numChannels,numSpec] = size(XTrain) %Витягує розмір XTrain за кожним виміром і присвоює їх змінним

Масштабуємо функції ступенем вікна і потім візьмемо логорифм. Щоб отримати дані з більш згладженим розподілом, візьмемо логарифм спектрограм за допомогою маленького зміщення.

```
epsil = 1e-6;
XTrain = log10(XTrain + epsil);
```

Виконаємо кроки вилучення ознак для валідаційного набору.

```
if ~isempty(ver("parallel"))
   pool = gcp;
   numPar = numpartitions(adsValidation,pool);
else
   numPar = 1;
end

parfor ii = 1:numPar
   subds = partition(adsValidation,numPar,ii);
   XValidation = zeros(numHops,numBands,1,numel(subds.Files));
   for idx = 1:numel(subds.Files)
      x = read(subds);
      xPadded = [zeros(floor((segmentSamples-size(x,1))/2),1);x;zeros(ceil((segmentSamples-size(x,1))/2),1)];
      XValidation(:,:,:,idx) = extract(aFE,xPadded);
```

```
    end
    XValidationC{ii} = XValidation;
end
XValidation = cat(4,XValidationC{:});
XValidation = log10(XValidation + epsil);
```

Ізолюємо мітки навчальних даних та валідації. Видалимо порожні категорії.

```
TTrain = removecats(adsTrain.Labels);
TValidation = removecats(adsValidation.Labels);
```

В.6. Візуалізація даних

Побудуємо криві та слухові спектрограми кількох навчальних зразків. Прослухаємо відповідні аудіокліпи.

specMin = min(XTrain,[],"all"); % обчислює мінімальне значення всіх елементів 4D матриці XTrain, яка є навчальними даними у вигляді Mel спектрограм

specMax = max(XTrain,[],"all"); % обчислює максимальне значення всіх елементів у XTrain

idx = randperm(numel(adsTrain.Files),3); %генерує випадкову перестановку цілих чисел від 1 до кількості аудіофайлів у adsTrain.Files, що є навчальним набором аудіофайлів. Перші три числа в перестановці присвоюються змінній idx.

figure(Units="normalized",Position=[0.2 0.2 0.6 0.6]); %створює нове вікно фігури з нормалізованою позицією та розміром.

% Чотири значення у векторі [0.2 0.2 0.6 0.6] задають лівий, нижній край, ширину та висоту фігури, відповідно, як частки від розміру екрану.

Цей код будує Mel спектрограми для трьох випадково вибраних прикладів з навчальної вибірки за допомогою функції subplot. Діапазон кольорової осі задається в межах [specMin, specMax] за допомогою функції caxis

```
for ii = 1:3
   [x,Fs] = audioread(adsTrain.Files{idx(ii)});
   subplot(2,3,ii)
   plot(x)
   axis tight
   title(string(adsTrain.Labels(idx(ii))))

   subplot(2,3,ii+3)
   spect = XTrain(:,:,1,idx(ii))';
   pcolor(spect)
   caxis([specMin specMax])
   shading flat
   sound(x,Fs)
   pause(2)
end
```

В.7. Додавання даних о фоновому шумі

Мережа повинна вміти не тільки розпізнавати різні вимовлені слова, але й визначати, чи містить вхідний сигнал тишу або фоновий шум.

Використаємо аудіофайли з папки *_background_* для створення зразків односекундних кліпів фонового шуму. Створимо однакову кількість фонових кліпів з кожного файлу фонового шуму. Також можна створити власні записи фонового шуму і додати їх до теки *_background_*. Перед обчисленням спектрограм функція перемасштабує кожен аудіокліп з коефіцієнтом, вибраним з логарифмічного рівномірного розподілу в діапазоні, заданому параметром volumeRange.

adsBkg = audioDatastore(fullfile(dataset,"background")) %створює об'єкт сховища даних аудіо з аудіофайлів у підтеці "background" каталогу "dataFolder"

numBkgClips = 4000; %задає бажану кількість фонових кліпів для генерації

if speedupExample %якщо "reduceDataset" - true

numBkgClips = numBkgClips/20; %зменшується кількість згенерованих фонових кліпів

end

volumeRange = log10([1e-4,1]); %задає діапазон підсилення, який буде застосовано до фонових кліпів у децибелах.

numBkgFiles = numel(adsBkg.Files); %повертає кількість фонових файлів у сховищі даних

numClipsPerFile = histcounts(1:numBkgClips,linspace(1,numBkgClips,numBkgFiles+1)); % розподіляє масив кліпів з кожного фонового файлу

% у numBkgFiles+1 рівновіддалених один від одного кошиків (bins) у межах від 1 до numBkgClips

%генерує масив кількості кліпів, які потрібно згенерувати з кожного фонового файлу, щоб досягти потрібної загальної кількості кліпів.

Xbkg = zeros(size(XTrain,1),size(XTrain,2),1,numBkgClips,"single"); %ініціалізує масив для зберігання фонових кліпів після вилучення елементів.

bkgAll = readall(adsBkg); %повертає всі дані зі сховища даних adsBkg

ind = 1;

for count = 1:numBkgFiles %обходить кожен фоновий файл.

bkg = bkgAll{count}; %витягує звукові дані з поточного файлу фонових аудіосигналів. випадковим чином вибираємо початковий індекс для кожного кліпу в поточному файлі

idxStart = randi(numel(bkg)-Fs,numClipsPerFile(count),1); % генерує вектор-стовпець розміром numClipsPerFile(count) x 1

% випадкових цілих чисел в інтервалі [1; numel(bkg)-Fs]

idxEnd = idxStart+Fs-1; %обчислює індекс закінчення для кожного кліпу на основі бажаної довжини кліпу з семплів "Fs".

gain = 10.^((volumeRange(2)-volumeRange(1))*rand(numClipsPerFile(count),1) + volumeRange(1)); %генерує випадковий коефіцієнт підсилення для кожного кліпу на основі вказаного діапазону.

for j = 1:numClipsPerFile(count) % обходить кожен кліп у поточному файлі

x = bkg(idxStart(j):idxEnd(j))*gain(j); %витягує аудіодані для поточного кліпу і застосовує випадковий коефіцієнт підсилення.

x = max(min(x,1),-1); %обмежує аудіосигнал значенням від -1 до 1

Xbkg(:,:,:,ind) = extract(aFE,x); % витягує ознаки для поточного кліпу і зберігає їх у масиві "Xbkg"

if mod(ind,1000)==0

progress = "Processed " + string(ind) + " background clips out of " + string(numBkgClips) %друкує повідомлення про прогрес після кожних 1000 оброблених кліпів

end

ind = ind + 1;

end

end

Xbkg = log10(Xbkg + epsil);

Розділимо спектрограми фонового шуму між навчальними, валідаційними та тестовими наборами.

numTrainBkg = floor(0.85*numBkgClips);%кількість фонових кліпів, використаних для навчання, що становить 85% від загальної кількості згенерованих фонових кліпів, numBkgClips.
numValidationBkg = floor(0.15*numBkgClips); %кількість фонових кліпів, використаних для перевірки, що становить 15% від загальної кількості згенерованих фонових кліпів.
XTrain(:,:,:,end+1:end+numTrainBkg) = Xbkg(:,:,:,1:numTrainBkg);%додає фонові аудіокліпи numTrainBkg у кінець навчальних даних (training data) XTrain.
TTrain(end+1:end+numTrainBkg) = "background"; %додає в кінець масиву YTrain мітки numTrainBkg зі значенням "background".
XValidation(:,:,:,end+1:end+numValidationBkg)=Xbkg(:,:,:,numTrainBkg+1:end);
TValidation(end+1:end+numValidationBkg) = "background";

Побудуємо графік розподілу різних міток класів у навчальній та валідаційній вибірках.

```
figure(Units="normalized",Position=[0.2 0.2 0.5 0.5])
tiledlayout(2,1)
nexttile
histogram(TTrain)
title("Training Label Distribution")
nexttile
histogram(TValidation)
title("Validation Label Distribution")
```

В.8. Створення архітектури НМ

Створимо мережеву архітектуру у вигляді масиву шарів. Використаємо згорткові шари та шари пакетної нормалізації, а також понизьмо дискретизацію карт ознак "просторово" (тобто за часом і частотою), за допомогою шарів максимального об'єднання (max-pooling). Додамо останній шар максимального об'єднання, який об'єднує вхідну карту ознак глобально глобально в залежності від часу. Це забезпечує (приблизну) часову інваріантність вхідних спектрограм, дозволяючи мережі виконувати однакову класифікацію незалежно від точного положення в часі. Глобальне об’єднання (Global pooling) також суттєво зменшує кількість параметрів у фінальному повноз’єднаному шарі. Щоб зменшити ймовірність запам'ятовування мережею специфічних особливостей навчальних даних, необхідно додати невелику кількість відсіяних даних на вхід останнього повноз’єднаного шару.

Змінна numF контролює кількість фільтрів у згорткових шарах. Щоб підвищити точність мережі, необхідно збільшити глибину мережі, додавши однакові блоки шарів згортки, пакетної нормалізації та ReLU. Також можна спробувати збільшити кількість згорткових фільтрів, збільшивши numF.

Щоб надати кожному класу рівну загальну вагу у втратах, використаємо ваги класів, обернено пропорційні кількості навчальних прикладів у кожному класі. При використанні оптимізатора Адама для навчання мережі алгоритм навчання не залежить від загальної нормалізації ваг класів.

classes = categories(Training_data);

classWeights = 1./countcats(Training_data); %отримаємо вектор ваг класів, який надає більшу вагу менш часто зустрічаючимся класам

classWeights = classWeights'/mean(classWeights); %ваги класів масштабуються за їх середнім значенням, що нормалізує їх таким чином, щоб їх сума дорівнювала кількості класів.

numClasses = numel(classes); %обчислює кількість класів шляхом підрахунку кількості унікальних категорій в YTrain

%Переконаємось, що область об'єднання часових інтервалів менша за вхідну спектрограму, щоб уникнути потенційної невідповідності розмірів під час навчання.

```
timePoolSize = ceil(numHops/10); %обчислення розміру області часового пулу
probability = 0.2;
numF = 20; % кількість фільтрів що використовуються у згорткових шарах мережі

layers = [
    imageInputLayer([numHops numBands])

    %1
    convolution2dLayer(3,numF,Padding="same")
    batchNormalizationLayer
    reluLayer

    maxPooling2dLayer(3,Stride=2,Padding="same")

    %2
    convolution2dLayer(3,2*numF,Padding="same")
    batchNormalizationLayer
    reluLayer

    maxPooling2dLayer(3,Stride=2,Padding="same")

    %3
    convolution2dLayer(3,2*numF,Padding="same")
    batchNormalizationLayer
    reluLayer
```

```
maxPooling2dLayer(3,Stride=2,Padding="same")

%4,5
convolution2dLayer(3,4*numF,Padding="same")
batchNormalizationLayer
reluLayer
convolution2dLayer(3,4*numF,Padding="same")
batchNormalizationLayer
reluLayer
maxPooling2dLayer(3,Stride=2,Padding="same")

%6,7
convolution2dLayer(3,6*numF,Padding="same")
batchNormalizationLayer
reluLayer
convolution2dLayer(3,6*numF,Padding="same")
batchNormalizationLayer
reluLayer
maxPooling2dLayer(3,Stride=2,Padding="same")

 %8,9
 convolution2dLayer(3,8*numF,Padding="same")
 batchNormalizationLayer
 reluLayer
 convolution2dLayer(3,8*numF,Padding="same")
 batchNormalizationLayer
 reluLayer
 maxPooling2dLayer(3,Stride=2,Padding="same")

 %10
```

```
 convolution2dLayer(3,10*numF,Padding="same")
 batchNormalizationLayer
 reluLayer
 maxPooling2dLayer([timePoolSize,1])
flattenLayer
dropoutLayer(probability)
fullyConnectedLayer(numClasses)
softmaxLayer
classificationLayer(Classes=classes,ClassWeights=classWeights)];
```

В.9. Навчання НМ

Вкажемо параметри навчання. Використовується оптимізатор Adam з розміром міні-партії 128. Навчання відбувається протягом 25 епох і зменшується швидкість навчання в 10 разів після 20 епох

```
miniBatchSize = 128;
validationFrequency = floor(numel(TTrain)/miniBatchSize);
options = trainingOptions("adam", ... %алгоритм оптимізації, який використовується для оновлення ваг нейронної мережі під час навчання
 InitialLearnRate=1e-3, ... %швидкість навчання, яка використовується для оновлення ваг на початку навчання
 MaxEpochs=25, ... %максимальна кількість епох (проходів через дані), на яку буде навчена мережа.
 MiniBatchSize=miniBatchSize, ... %розмір міні-партій
 Shuffle="every-epoch", ... %опція вказує, що навчальні дані повинні бути випадковим чином перемішані на початку кожної епохи
 Plots="training-progress", ... %параметр вказує, що під час навчання мають відображатися графіки прогресу навчання
```

Verbose=false, ... %параметр визначає, чи відображати інформацію про хід навчання у командному вікні

ValidationData={XValidation,TValidation}, ... %параметр визначає дані перевірки, які використовуватимуться під час тренування

ValidationFrequency=validationFrequency, ... %Частота перевірки - параметр визначає, як часто оцінювати дані валідації під час навчання (у вигляді міні-партій)

LearnRateSchedule="piecewise", ... %параметр визначає графік швидкості навчання. 'piecewise' означає, що швидкість навчання буде оновлюватися в певні проміжки часу

LearnRateDropFactor=0.1, ...%параметр визначає коефіцієнт, на який зменшуватиметься швидкість навчання, коли спрацьовуватиме LearnRateSchedule

LearnRateDropPeriod=20); %опція визначає, як часто зменшувати швидкість навчання, коли спрацьовує LearnRateSchedule

trainedNet = trainNetwork(XTrain,TTrain,layers,options); %навчання нейронної мережі з вказаними навчальними даними, мітками, шарами та опціями

В.10. Оцінка навченої мережі

Обчислимо остаточну точність мережі на навчальному наборі даних (без доповнення даних) та валідаційному наборах. Мережа дуже точна на цьому наборі даних. Однак навчальні, валідаційні та тестові дані мають схожий розподіл, який не обов'язково відображає реальні умови. Це обмеження особливо стосується категорії unknown.

% if speedupExample

% load("commandNet.mat","trainedNet"); %завантажує модель "commandNet.mat" у змінну trainedNet

% end

YValPred = classify(trainedNet,XValidation); %класифікує валідаційний набір даних за допомогою навченої мережі trainedNet і зберігає передбачені мітки в YValPred

validationError = mean(YValPred ~= TValidation); %обчислює частоту помилок валідації, порівнюючи передбачені мітки YValPred з фактичними мітками TValidation

YTrainPred = classify(trainedNet,XTrain); %класифікує навчальний набір даних, використовуючи ту саму навчену мережу trainedNet, і зберігає передбачені мітки у змінній YTrainPred

trainError = mean(YTrainPred ~= TTrain); %обчислює частоту помилок навчання, порівнюючи передбачені мітки YTrainPred з фактичними мітками TTrain

disp("Training error: " + trainError*100 + "%")

disp("Validation error: " + validationError*100 + "%")

Побудуємо матрицю помилок (confusion matrix).

figure(Units="normalized",Position=[0.2 0.2 0.5 0.5]);

cm = confusionchart(TValidation,YValPred, ... %створюється матрична діаграма помилок

Title="Confusion Matrix for Validation Data", ... %Властивість Title використовується для встановлення заголовка діаграми.

ColumnSummary="column-normalized", ... %Властивості ColumnSummary використовуються для нормалізації рядків матриці помилок

RowSummary="row-normalized"); %Властивості RowSummary використовуються для нормалізації стовпців матриці помилок

sortClasses(cm,[commands,"unknown","background"]) % сортування класів у матриці помилок

Для оцінки якості роботи алгоритму на кожному з класів окремо розрахуємо метрики *precision, recall та F1-score*.

Властивість *NormalizedValues* об'єкта *confusionchart* містить значення нормалізованої матриці помилок.

Застосувавши функцію sum до cm.NormalizedValues з відповідним аргументом розмірності (2 для рядків і 1 для стовпців), ви можете обчислити суму для кожного рядка і стовпця відповідно.

```
diagonal = diag(cm.NormalizedValues) %Виберемо основну діагональ матриці
sum_of_rows = sum(cm.NormalizedValues, 2) %Розрахуємо суму значень для кожного ряду
precision = diagonal ./ sum_of_rows % Точність (precision)
overall_precision = mean(precision) %Загальна точність мережі
sum_of_columns = sum(cm.NormalizedValues, 1) %Розрахуємо суму значень для кожного стовпця
recall = diagonal ./ sum_of_columns' %Відновлюваність (recall)
overall_recall = mean(recall) % Загальна відновлюваність
f1_score = 2*((overall_precision*overall_recall)/(overall_precision+overall_recall)) %Оцінка F1
```

Обчислимо загальний розмір мережі в кілобайтах і протестуйте швидкість її передбачення при використанні центрального процесора. Час передбачення - це час класифікації одного вхідного зображення. Для класифікації потокового аудіо найбільш важливим є час розпізнавання одного зображення

Цей код вимірює час, необхідний для класифікації одного зображення за допомогою trainedNet на центральному процесорі.

```
info = whos("trainedNet");

disp("Network size: " + info.bytes/1024 + " kB")
```

```
time = zeros(100,1);
%Цикл виконується 100 разів, і на кожній ітерації генерується випадкове вхідне зображення x, яке пропускається через мережу за допомогою функції classify.
for ii = 1:100
   x = randn([numHops,numBands]);
   tic %Час виконання кожної ітерації
   [YPredicted,probs] = classify(trainedNet,x,ExecutionEnvironment="cpu");
   time(ii) = toc; % Час виконання кожної ітерації зберігається в масиві часу.
end

% середній час виконання обчислюється як середнє значення елементів масиву часу, починаючи з 11-ї ітерації
% (щоб виключити будь-які накладні витрати на ініціалізацію). Результат відображається в мілісекундах.
disp("Single-image prediction time on CPU: " + mean(time(11:end))*1000 + " ms")
```