Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)
Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

бакалавр
(назва освітнього ступеня)

на тему: Розробка моделі акустичного придушення ехо в активній аудіосистемі

Виконав: студент IV курсу, групи СН-42
спеціальності 122 Комп’ютерні науки
(шифр і назва спеціальності)

(підпис) Світлик П.А. (прізвище та ініціали)

Керівник (підпис) Липак Г.І. (прізвище та ініціали)

Нормоконтроль (підпис) Шимчук Г.В. (прізвище та ініціали)

Завідувач кафедри (підпис) Боднарчук І.О. (прізвище та ініціали)

Рецензент (підпис) Тиш Є. В. (прізвище та ініціали)

Тернопіль - 2025

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)
Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Боднарчук І.О.
(підпис) (прізвище та ініціали)
«__» _____ 2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня Бакалавр
(назва освітнього ступеня)
за спеціальністю 122 Комп'ютерні науки
(шифр і назва спеціальності)
Студенту Світлику Павлу Андрійовичу
(прізвище, ім'я, по батькові)
1. Тема роботи Розробка моделі акустичного придушення ехо в активній аудіосистемі

Керівник роботи Липак Галина Ігорівна, к.соц. ком., доц.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)
Затверджені наказом ректора від «07» 05 2025 року № 4/7-444
2. Термін подання студентом завершеної роботи 16.06.2025 р.
3. Вихідні дані до роботи наукові літературні джерела

4. Зміст роботи (перелік питань, які потрібно розробити)
Вступ
1. Аналіз предметної області.
2. Теоретична частина
3. Практична частина
4. Безпека життєдіяльності, основи охорони праці
Висновки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)
1. Титулка. 2. Мета, задачі дослідження. 3. Схема запропонованої моделі виділення корисного сигналу s(n). 4. Метрики якості моделі. 5. Вхідні дані для моделі. 6. Вихід моделі. 7, 8. Структура пропонованої моделі (навчання, тестування). 9. Технології та засоби для реалізації моделі. 10. Оцінка ефективності моделі BLSTM+k-Means. 11. Порівняння ефективності моделей з двома розмовами. 12. Порівняння ефективності моделей з однією розмовою. 13. Висновки. Основні результати проведеного дослідження

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Безпека життєдіяльності, основи охорони праці	Сенчишин В.С., доц. каф. МТ		

7. Дата видачі завдання 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	08.05 – 09.05.25	*Виконано*
2.	Підбір джерел про моделі акустичного ехопридушення	10.05 – 14.05.25	*Виконано*
3.	Опрацювання джерел про моделі акустичного		
	ехопридушення	15.05 – 19.05.25	*Виконано*
4.	Виконання дослідження щодо розробки моделі		
	акустичного ехопридушення в активній аудіосистемі	20.05 – 23.05.25	*Виконано*
5	Розроблення програмного коду	24.05 –27.05.25	*Виконано*
6.	Оформлення розділу «Аналіз предметної області»	26.05 – 29.05.25	*Виконано*
7.	Оформлення розділу «Теоретична частина»	30.05 – 02.06.25	*Виконано*
8.	Оформлення розділу «Практична частина»	03.06 – 06.06.25	*Виконано*
9.	Виконання завдання до підрозділу «Безпека		
	життєдіяльності, основи хорони праці»	31.05 – 03.06.25	*Виконано*
10.	Оформлення кваліфікаційної роботи	06.06 – 09.06.25	*Виконано*
11.	Нормоконтроль	10.06 – 13.06.25	*Виконано*
12.	Перевірка на плагіат	14.06 – 17.06.25	*Виконано*
13.	Попередній захист кваліфікаційної роботи	18.06 – 20.06.25	*Виконано*
14.	Захист кваліфікаційної роботи	23.06.25	

Студент _____ Світлик П.А.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Липак Г.І.
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Розробка моделі акустичного придушення ехо в активній аудіосистемі // Світлик Павло Андрійович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем та програмної інженерії, кафедра комп'ютерних наук, група СН–42 // Тернопіль, 2025 // с. – 57, рис. – 28, табл. – 3, слайдів – 12, бібліогр. – 34.

Ключові слова: BLSTM, алгоритми кластеризації, ехо-сигнал, ехопридушення, нейронна мережа, перетворення Фур'є.

Кваліфікаційна робота присвячена розробці алгоритму та моделі на його базі, котра здатна придушувати ехосигнали.

У першому розділі було поставлено та описано завдання на виконання дослідження, також були проведені оглядові роботи, пов'язані із тематикою дослідженням.

Другий розділ містить теоретичну частину, в якій вивчалися і аналізувалися алгоритми та методи, використані в роботі. Розроблено алгоритм на основі BLSTM, виходом якої є ідеальна бінарна маска. Ключовою особливістю запропонованого алгоритму є використання методів кластеризації (EM, Mean-Shift, k-Means) на виході нейронної мережі.

У третьому розділі було описано та реалізовано запропоновану модель BLSTM+clustering. Проведено порівняння алгоритмів на сигналах бази даних TIMIT на основі загальноприйнятих метрик у обробці мовлення: ERLE, STOI, PESQ. Наведено результати експериментів та порівняння ефективності моделей. Показано, що використання кластеризації k-Means покращує роботу моделі BLSTM.

У четвертому розділі висвітлено важливі питання охорони праці та безпеки життєдіяльності.

ANNOTATION

Development of an acoustic echo suppression model in an active audio system // Svitlyk Pavlo // Ternopil Ivan Pul'uj National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science // Ternopil, 2025 // P. - 57, Fig. - 28, Table - 3, Slide - 12, References - 34.

Keywords: BLSTM, clustering algorithms, echo signal, echo suppression, neural network, Fourier transform

The thesis deals with the development of an algorithm and a system based on it, which is capable of suppressing echo signals.

In the first chapter, the tasks for the research were set and described, and review works related to the research topic were also carried out.

The second chapter contains the theoretical part, in which the algorithms and methods used in the work were studied and analyzed. An algorithm based on BLSTM was developed, the output of which is an ideal binary mask. The key feature of the proposed algorithm is the use of clustering methods (EM, Mean-Shift, k-Means) at the output of the neural network.

In the third chapter, the proposed BLSTM+clustering model was described and implemented. A comparison of algorithms on signals from the TIMIT database was carried out based on generally accepted metrics in speech processing: ERLE, STOI, PESQ. The results of experiments and a comparison of the effectiveness of the models are presented. It is shown that the use of k-Means clustering improves the performance of the BLSTM model.

The fourth chapter highlights important issues of occupational health and safety.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ СКОРОЧЕНЬ І ТЕРМІНІВ

BLSTM (Bidirectional Long Short-Term Memory) – двонаправлена рекурентна нейронна мережа.

EM (Expectation-maximization) – очікування-максимізації.

ERLE (Echo Return Loss Enhancement) – рівень ослаблення ехо-сигналу.

IBM (Ideal Binary Mask)– ідеальна бінарна маска.

ISM (Image Source Method) – метод симетричного відображенням першоджерела по відношенню до зустрічної границі.

ISTFT (Inverse Short Time Fourier Transform) – зворотне короткочасне перетворення Фур'є.

LMS (Least Mean Square) – алгоритм найменших середніх квадратів

LSTM (Long short Term memory) – довга короткочасна пам’ять, є архітектурою рекурентних нейронних мереж.

PESQ ((Perceptual Evaluation of Speech Quality) – перцептивна оцінка якості мовлення (метрика).

RIR (Room Impulse Responses) – характеристика обробки аудіосигналу, яка передбачає захоплення та аналіз акустичних характеристик кімнати.

RLS (Recursive Least Squares) – алгоритм рекурсивних найменших квадратів.

SER (Signal-to-Echo Ratio) – відношення сигнал/ехо.

STFT (Short Time Fourier Transform) – короткочасне перетворення Фур'є.

STOI (Short Time Objective Intelligibility) – короткочасна об’єктивна зрозумілість мовлення (метрика).

TIMIT - збірник фонематично та лексично транскрипованого мовлення носіїв американської англійської мови різної статі та діалектів.

ЗМІСТ

ВСТУП

Актуальність теми. Алгоритми відновлення мовного сигналу, спотвореного адитивним некорельованим шумом, у разі, коли доступний тільки зашумлений сигнал, широко застосовуються в різних областях цифрової обробки мовних сигналів, таких як розпізнавання мовлення, розпізнавання особи, котра говорить, детектування мовленнєвої активності, поліпшення якості та розбірливості мовленнєвих сигналів [1].

З розвитком ефективних методів машинного навчання стала вельми поширеною практика отримувати алгоритми придушення шуму з урахуванням глибинних нейронних мереж [2-4].

Одними з найбільш використовуваних методів шумозаглушення є методи, засновані на оцінці частотно-часових масок [5]. Наприклад, у роботах [6, 7] як цільовий вихід нейромережевої моделі є IBM.

Проте недостатньо уваги приділено використанню BLSTM, методів кластеризації (зокрема EM, Mean-Shift, k-Means) на виході моделі та підрахунку і порівнянню метрик у обробці мовлення (ERLE, STOI, PESQ) для забезпечення ехопридушення в кімнаті.

Мета роботи – побудувати модель придушення акустичного еха в активній аудіосистемі на основі нейромереж та методів кластеризації.

Для досягнення мети виділено ряд завдань:

– провести аналіз існуючих методів та алгоритмів придушення акустичного еха;

– провести моделювання та підготувати вихідні дані;

– розробити технології реалізації моделі;

– навчити рекурентну нейронну мережу BLSTM з використанням середовища Python у хмарі Google Colab;

– оцінити ефективність моделі;

– порівняти ефективність запропонованої моделі при використанні алгоритмів кластеризації (K-Means, Mean-Shift, EM), які застосовуються до мережевих вихідних вибірок.

Практичне значення. Запропонований алгоритм та модель на його основі може бути використана як складова системи для забезпечення придушення ехо-сигналу.

РОЗДІЛ 1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Постановка завдання на проектування

Розглянута модель наведена на рис. 1.1. Сигнал мікрофона *y(n)* складається з сигналу на ближньому кінці *s(n)*, еха *d(n)* і фонового шуму *v(n):*

$$y(n) = d(n) + s(n) + v(n) \tag{1.1}$$

Для простоти у цій роботі вважатимемо *v(n)=0.*

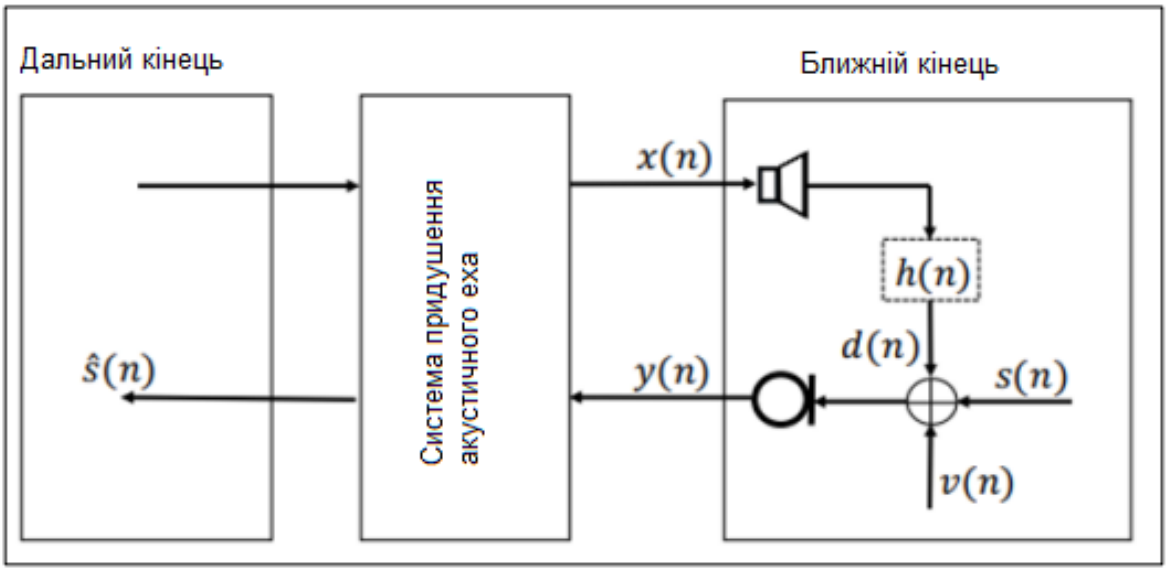

Рисунок 1.1 – Адитивна модель акустичного ехо

Ехо-сигнал *d(n)* формується за рахунок відображення сигналу з далекого кінця *x(n)* від стінок кімнати та моделюється шляхом згортки *x(n)* з імпульсною характеристикою *h(n)* приміщення RIR.

Наше завдання - виділити із суміші *y(n)* корисний сигнал *s(n),* прибравши небажану перешкоду *d(n).*

На рис. 1.2 представлена схема запропонованої моделі на вирішення цієї задачі. З сигналу мікрофона *y(n)* за допомогою STFT витягуються ознаки, які є вхідними даними для BLSTM. Виходом нейронної мережі є IBM, яка часто

використовується як мета в задачі поділу мовлення від перешкод.

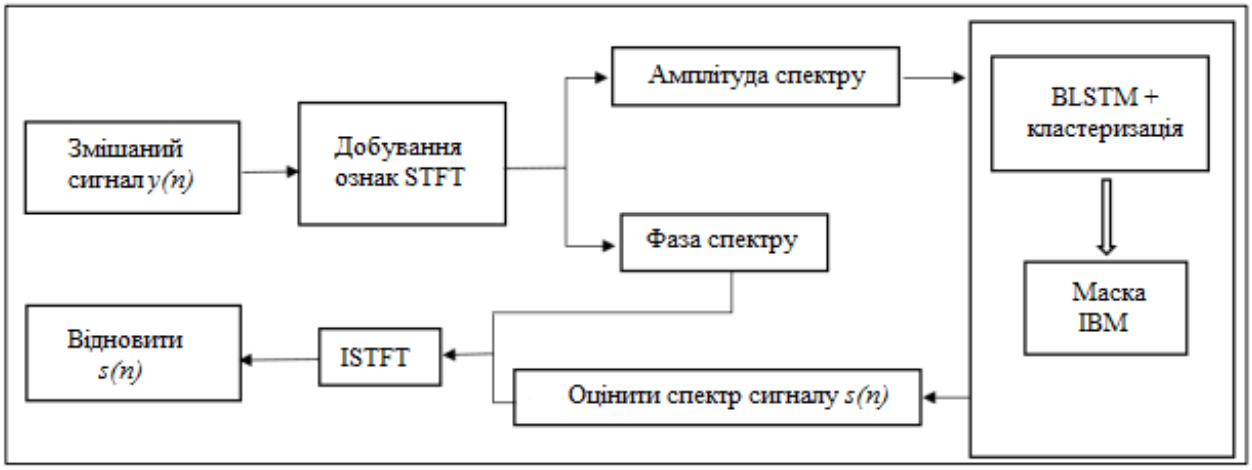

Рисунок 1.2 – Схема запропонованої моделі виділення сигналу *s(n)*

Використовуючи маску IBM, можна оцінити спектр сигналу ближнього кінця і, за допомогою ISTFT, відновити *s(n)*.

1.2 Огляд аналогів

1.2.1 Огляд пов'язаних робіт

Традиційно завдання акустичного ехоподавлення вирішується за рахунок адаптації акустичної імпульсної характеристики між гучномовцем та мікрофоном з використанням фільтра з кінцевою імпульсною характеристикою [8]. Одним із найбільш широко використовуваних адаптивних алгоритмів є NLMS (Normalized Least Mean Square) [9], що має хорошу надійність при низькій складності. Однак він, як і всі адаптивні алгоритми [8], має недолік можливої витратності через кореляцію між корисним сигналом і ехом. Ця кореляція має місце, наприклад, під час так званої подвійної розмови і більшість алгоритмів будується на спробі сторонніми засобами розпізнати подвійне мовлення в конкретний момент часу і призупинити навчання адаптивного фільтра [10]. За допомогою такого методу можна запобігти розбіжності фільтра, проте навчання значно сповільнюється.

Сигнал, що надходить на мікрофон, містить не тільки відлуння і мовлення

на ближньому кінці, а й фоновий шум, з яким система акустичного ехоподавлення сама по собі не впорається. Для придушення фонового шуму та залишкових ехо-сигналів, які існують на виході системи, зазвичай використовують пост-фільтри. Для прикладу, у роботі [11] автори об'єднали адаптивний алгоритм із методом придушення шуму з урахуванням короткочасного спектрального згасання і отримали високий рівень видалення луна за наявності фонового шуму.

Іншим варіантом усунення залишкового еха є використання потужних моделей глибинного навчання [12], котрі здатні моделювати складні нелінійні спотворення, що вносяться підсилювачами потужності та гучномовцями. Такі моделі є все більш популярнішою альтернативою нелінійному моделюванню систем акустичного ехоподавлення.

Так, автори роботи [13] змоделювали нелінійну систему як модель Хаммерштейна і використовували повнозв'язкову нейронну мережу з двома шарами, за якою слідує адаптивний фільтр для ідентифікації параметрів моделі. У роботі [2] використано глибинну нейронну мережу для оцінки посилення залишкового шуму як на дальньому кінці сигналу, так і на виході системи [14], щоб видалити нелінійні компоненти ехо-сигналу. Глибинне навчання показало великий потенціал для поділу мовлення [15, 16] . Здатність рекурентних нейронних мереж RNN моделювати функції, що змінюються в часі, може відігравати важливу роль у вирішенні проблем акустичного ехоподавлення.

LSTM [17] - варіант RNN, розроблений для вирішення проблеми зникнення градієнта, властивої традиційним RNN, вже показав хорошу продуктивність у завданнях розпізнавання та покращення мовлення в галасливих умовах [18, 19]. Робота [20] також відзначає найкращу якість моделі, що використовує LSTM у порівнянні з глибокими повнозв'язковиим нейронними мережеми.

1.2.2 Відомі підходи

Основною вимогою до алгоритму ехоподавлення є самостійна адаптація до конкретної звукової системи. У реальному застосуванні конфігурація

відлуння шляхів невідома і може активно змінюватися в часі. Для таких завдань існує клас алгоритмів в галузі цифрової обробки сигналів, що отримав назву адаптивних фільтрів. Фільтром у сенсі слова називається система, яка виділяє чи поліпшує певну інформацію, що міститься у сигналі.

Цифрові адаптивні фільтри дуже популярні нині завдяки їх високій ефективності при обробці сигналів та навчальної природі. Навчальність досягається через існування негативного зворотного зв'язку за сигналом помилки. Зазвичай адаптивний фільтр має в складі певний алгоритм автопідлаштовування, який змінює коефіцієнти фільтра на підставі сигналу помилки. Стан фільтра визначається через деякий вектор коефіцієнтів *w(n).*

Розроблено велику кількість адаптивних фільтрів, всі вони мають різні переваги та недоліки. Важливо відзначити, що сфери застосування таких фільтрів так само значно різняться.

Наведемо схематичне представлення звукової системи, у якій адаптивний фільтр видаляє зворотне ехо. На рис. 1.3 видно, що цифровий сигнал *x(n)* програється динаміками і, проходячи за ехо-шляхом, піддається впливу невідомого перетворення *w(n).* Це відбувається під впливом акустики конкретного приміщення. Надалі цей звуковий сигнал поєднується з корисним сигналом розмови і шумом *v (n).* У такому вигляді сигнал потрапляє у мікрофон і відбувається його оцифрування у відповідному звуковому пристрої. Цифровий сигнал позначимо як *d(n).*

Процес роботи адаптивного фільтра є ітераційним, на кожній з ітерацій відбуваються дві основні дії: фільтрація ехо сигналу і адаптація коефіцієнтів фільтра w(n) з допомогою алгоритму автоналаштовування.

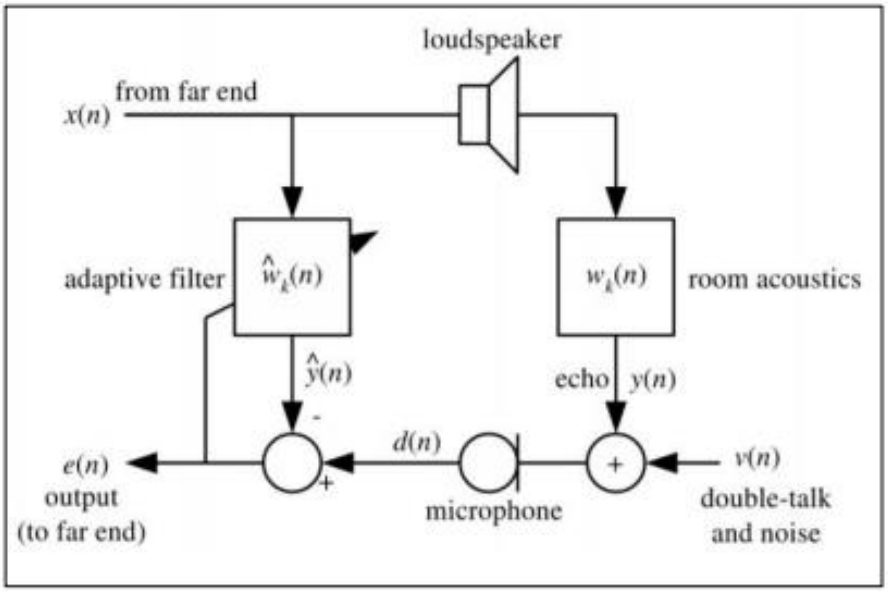

Рисунок 1.3 – Схема ехоподавлення адаптивним фільтром

Перший етап роботи фільтра на черговій ітерації - передбачення еха сигналу $y(n)$ у даний момент часу (відлік під номером n). Це передбачення засноване на збереженій історії $x(n)$ сигналу, що корелює з ехом. У даному випадку це сигнал з динаміків. Довжина цього вектора позначається як N і називається довжиною фільтра.

$$x(n) = [x(n), x(n-1), \ldots., x(n-N+1)]^T \tag{1.2}$$

де позначення $(.)^T$ є операцією транспонування.

Віднімаючи передбачене ехо із записаного мікрофоном сигналу $d(n)$, отримуємо сигнал помилки на даному відліку:

$$e(n) = d(n) - y(n) = v(n) + \Delta y(n) \tag{1.3}$$

Цей процес є фільтрацією небажаного сигналу. Обчислений сигнал помилки містить як корисний сигнал $v(n)$, так і $\Delta y(n) = \hat{y}(n) - y(n)$, тобто різницю між справжнім ехом та обчисленим фільтром.

Звернімо увагу, що сигнал помилки є результатом роботи фільтра. Другий

етап роботи фільтра полягає в адаптації його коефіцієнтів на підставі алгоритму автопідлаштовування. Для цього сигнал помилки $e(n)$ надходить на вхід алгоритму. Це є негативним зворотним зв'язком, оскільки основним завданням фільтра з математичної погляду є мінімізація $|\Delta y(n)|$, отже, і помилки $|e(n)|$. Виразимо це через середньоквадратичну помилку $E\{e^2(n)\}$, яка є критерієм ефективності фільтра.

Існує два основні адаптивні алгоритми:

– LMS;

– RLS.

РОЗДІЛ 2. ТЕОРЕТИЧНА ЧАСТИНА

2.1 Аналіз акустичного еха

Розглянемо джерело звуку або подію, розташовані на *s*, а приймач або мікрофон на *r*. Крім того, s і *r* - вектори положення в тривимірному просторі, вони також використовуються для позначення фактичного джерела і мікрофон на цих позиціях.

Коли *s* видає звук у момент часу T_0 усередині кімнати, частина звуку прямує прямо до приймача *r* і надходить вчасно $t = T_0 + \tau_0$. Тут τ_0 - час затримки прямого звуку до приймача і визначається як:

$$\tau_0 = \frac{\|s - r\|_2}{c} \tag{2.1}$$

Тут *c* – швидкість звуку. Крім того, інші частини звуку спочатку відбиватимуться від поверхонь у кімнаті до досягнення *r* в $t = T_0 + \tau_i$; $\tau_i > \tau_0$. Через нерівність трикутника відбитий звук завжди буде приходити до приймача пізніше, ніж прямий звук. Звук, який одного разу відбився, називається першим відображенням порядку. Відбиття другого порядку двічі відбивались від поверхонь.

На рис. 2.1 показаний прямий звук та одне ехо першого та другого порядку. Прямий звук позначений синьою пунктирною лінією, відображення першого та другого порядку з чорними та помаранчевими пунктирними лініями відповідно.

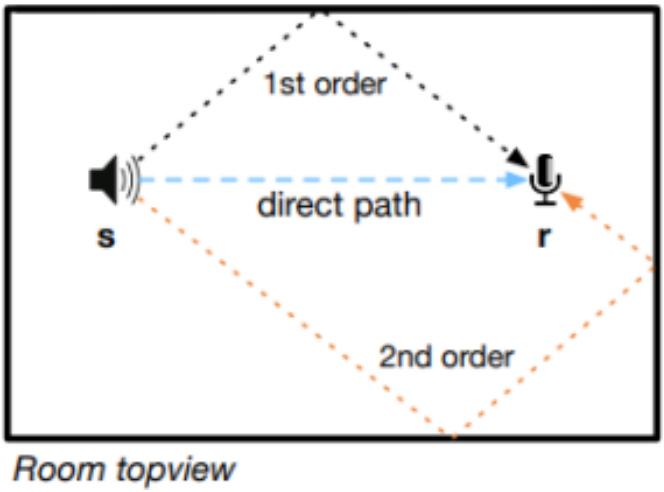

Рисунок 2.1 – Джерело звуку *s* і приймач *r* розташовані всередині кімнати

Коли джерело збуджує ідеальний імпульс, всі відображення, які досягають приймача, разом становлять імпульсну характеристику приміщення RIR, яка в ідеалі є послідовністю імпульсів, кожен з яких відповідає одному exo, може бути виражена як:

$$h(t) = \sum_i \alpha_i \delta(t - \tau_i) \tag{2.2}$$

Тут a_i коефіцієнт відбиття, пов'язаний з i-м відбиттям. Коефіцієнт відбиття послаблює енергію цього конкретного відбиття і залежить від типу матеріалу поверхні, порядку відбиття та відстані, котру проходить звук.

На рис. 2.2 показано ідеальну RIR, засновану на ситуації з рис. 2.1. Ми бачимо з рнс. 2.2, що імпульси відповідають відбиттям. Відбиття другого порядку має меншу енергію, ніж відбиття першого порядку, оскільки воно двічі відбивався від поверхні.

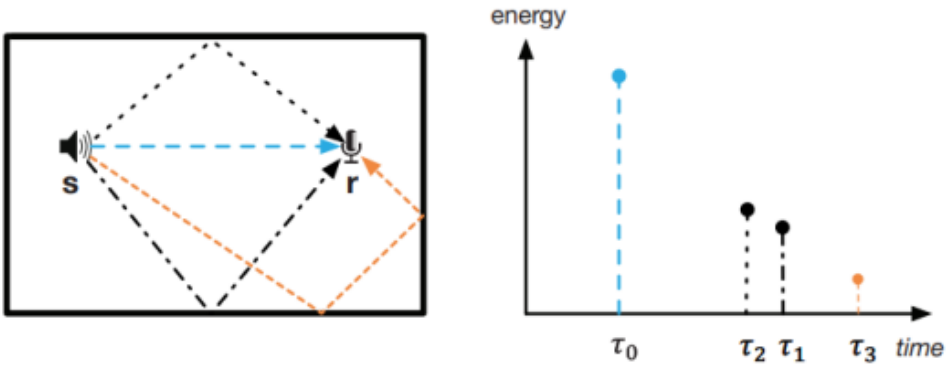

Рисунок 2.2 – Ідеальна RIR

На будь-який звук, що видається у кімнаті, впливатиме RIR. У кожному місці в кімнаті сигнал *y(t)*, який досягне приймача, буде результатом згортки (convolution) оригіналу звукового сигналу *x(t)* з *h(t)*:

$$y(t) = x * h(t) = \int x(t)h(t-\tau)d\tau = \int \alpha(\tau)x(t-\tau)d\tau \qquad (2.3)$$

Зверніть увагу, що *h(t)* є специфічним для кожного розташування *r* і *s*, оскільки час відбиття буде іншим.

2.2 Рекурентна LSTM нейромережа

Системам глибинного навчання вдалося досягти дуже хороших результатів для великого діапазону завдань. Високе навчання на величезних навчальних вибірках містить сотні тисяч вже розмічених примірників даних.

Причина, через яку системи глибинного навчання для одержання задовільного результату потребують великих навчальних вибірок, полягає у ітеративній, властиво, природі навчання. Ідея в тому, що модель раз за разом обновляє множини ваг на власне своїх синапсах для кожного примірника навчання, застосовуючи методи локальної оптимізації, фактично найбільш поширеним є стохастичний градієнтний спуск.

Висока результативність рекурентних нейронних мереж зумовлена їх властивістю використовувати наступні кроки інформацію з попередніх, тобто

мати деяку «пам'ять», на відміну нейронних мереж прямого поширення.

Однак у простих рекурентних нейронних мереж є суттєвий недолік - при великій кількості рекурентних ітерацій інформація з попередніх поступово «розмивається». Це відбувається через пропуск сигналу всередині комірки через функцію активації (рис. 2.3), нормалізуючи вихідний сигнал, для приведення його до деякого діапазону значень. Для вирішення цього обмеження розроблена модифікація LSTM має більш складну структуру комірки.

LSTM володіє властивістю видалення чи додавання інформації до клітинного стану, але така властивість старанно регулюється структурами, котрі носять назву вентилів (gates). Вони є способом часткового пропускання інформації. Вентилі містять сигмоїдний шар мережі і операції множення поточково (pointwise multiplication).

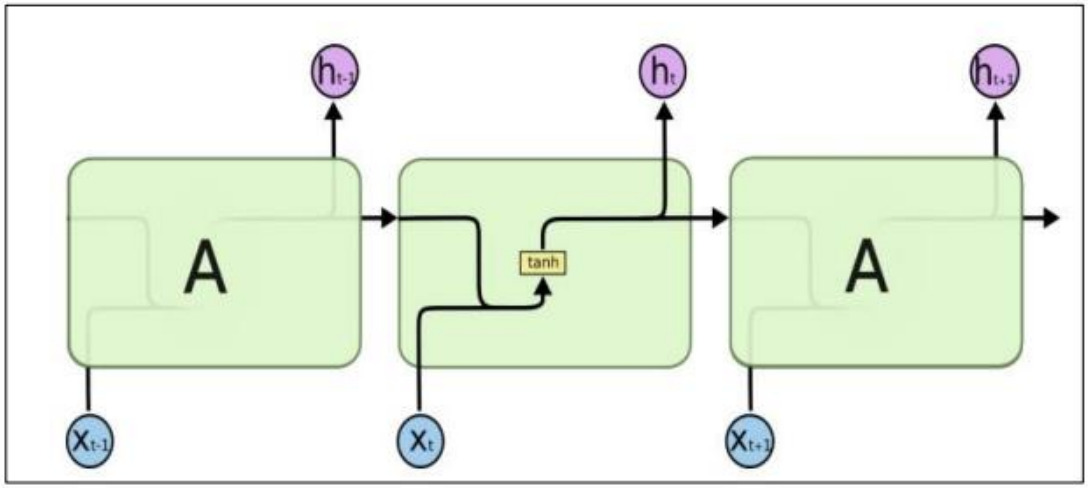

Рисунок 2.3 – Загальне представлення архітектури традиційної рекурентної нейромережі (X - вхідний сигнал, h - вихідний сигнал, A – комірка рекурентної нейромережі, tanh – функція активації гіперболічного тангенсу)

Сигмоїдний шар надсилає на вихід величини від нуля та одиниці, як підсумок вказуючи, наскільки кожен екземпляр може бути проведеним через вентиль. Тут 0 – «не проводити нічого», 1 – «все проводити».

LSTM -блоки мають три такі вентилі, котрі застосовуються з метою забезпечення контролювання інформаційних потоків на входах та виходах властиво пам'яті цих блоків. Такі вентилі втілені як функції логістики для

знаходження значення на діапазоні [0, 1]. Власне помноження такого значення застосовується для часткового допущення або заборонення інформаційного потоку всередину та назовні пам'яті [21].

У завданнях з видаленням ехо-сигналів широко використовується двонаправлена нейронна мережа. Такі мережі характеризуються здатністю передачі майбутніх станів на попередні шари, тоді як в односпрямованих рекурентних нейронних мережах на це накладено обмеження архітектурою.

2.3 Алгоритм кластеризації K-Means

Прискорений прогрес технологій останнім часом сприяє значному збільшенню обсягу створюваної та збереженої інформації в таких галузях, як освіта, інженерія, навчання, медицина та торгівля. Відповідно, існує зацікавленість у отриманні цінної інформації, яку можна добути з цих величезних обсягів інформації, щоб допомогти зробити найкращий вибір та зрозуміти ідею інформації. Кластеризація - один із фундаментальних методів для розуміння лежачого в основі характеру та структури даних. Кластеризація, або кластерний аналіз, є завданням навчання, яке виконується без будь-якого контролю з боку людини. Цей метод часто використовувався з великою ефективністю при аналізі даних і служить для спостереження та виявлення цікавих, корисних чи бажаних закономірностей у зазначених даних. Техніка кластеризації функціонує, виконуючи структурований поділ задіяних даних на аналогічні об'єкти на основі їх характеристик, які вона ідентифікує. Цей процес призводить до формування груп і кожна сформована група називається кластером. Один зазначений кластер складається з об'єктів даних, які мають схожість з іншими об'єктами, знайденими в тому ж кластері, і нагадують відмінності при порівнянні з об'єктами, ідентифікованими з даних, які в даний час існують в інших кластерах [22].

Кластеризація - це процес вибірки або набору точок даних на декілька груп так, щоб точки даних групи були більш схожими, ніж точки даних в інших групах. Процес кластеризації дуже важливий у різних аспектах аналізу даних,

оскільки він є використанням угруповання об'єктів, присутніх у даних, на основі їх атрибутів у пакеті необроблених даних [22].

Одним з найбільш широко застосовуваних на сьогоднішній день алгоритмів кластеризації є K-Means через простоту досягнення його результатів та реалізації. Термін «K -Means» вперше був використаний Джеймсом Маккуїном в 1967 р. [22] . K -Means - широко використовується як алгоритм інтелектуального аналізу даних, який можна використовувати для кластеризації даних шляхом аналізу та порівняння числових значень із самих даних. Це добре відомий метод поділу, що є ітеративним процесом. У цьому випадку об'єкти кластеризуються як такі, що належать до однієї з груп, а потім розподіляються за найближчим кластером. Кластер означає об'єднання елементів, що становлять подібність, шляхом простого поділу на групи.

Кроки алгоритму K-Means є такими.

1. Ініціалізація: спочатку виберемо кількість класів.
2. Ініціалізувати відповідні центральні точки даних як центроїди.
3. Призначення центроїду: кожна точка кластеризується шляхом обчислення відстані. Ці точки потім призначаються найближчому центроїду кластера.
4. Корекція центроїдів: наступним кроком є обчислення центроїдів новоствореного кластера.
5. Порівняння різниць: знову обчисліть центр групи, взявши середнє значення всіх векторів у групі.
6. Повторюємо ці кроки для заданої кількості ітерацій або до тих пір, поки центри груп не сильно змінюватимуться між ітераціями.

Вибір розрахунку відстані є важливим кроком. Два основних методи, що використовуються для розрахунку відстаней:

– Евклідова відстань (Euclidean distance);
– Манхеттенська відстань (Manhattan distance).

У цьому дослідженні використовується метод Евклідової відстані через його простоту розрахунку відстані подібності. Це стосується відстані між двома точками даних або відстані між точкою даних та центром кластера. Евклідова

відстань власне між двома точками, x і y, розмірністю k розраховується так [22]:

$$Dist_{xy} = \sqrt{\sum_{k=1}^{m}(x_{ik} - y_{ik})^2} \quad (2.4)$$

Евклідова відстань або евклідова метрика – це звичайна відстань між властиво двома точками, яку можна виміряти лінійкою. Це відстань по прямій лінії м між двома точками.

Блок-схема алгоритму K-Means наведена на рис. 2.4.

Рисунок 2.4 – Блок-схема алгоритму K-Means

K-Means надзвичайно чутливий до ініціалізації, а погана ініціалізація може навіть призвести до поганої швидкості збіжності та поганої загальної кластеризації. Отже, найкраще ініціалізувати дані випадковим чином.

2.4 Короткочасне перетворення Фур'є

Спектр мовних сигналів відіграє важливу роль у мовленні, за допомогою якої можуть бути отримані деякі нетривіальні мовленнєві ознаки. Стаціонарні сигнали відносяться до сигналів, закон розподілу яких не змінюється з часом, а нестаціонарні сигнали - це сигнали, що змінюються у часі. Змішані мовленнєві сигнали відносяться до нестаціонарних сигналів, а мовленнєві сигнали мають ознаки нестаціонарних.

Перетворення Фур'є FT (Fourier Transform) широко застосовується у науковій і технічній сферах. Для прикладу, воно знаходить застосування при вирішенні рівнянь для потоку тепла, для дифракції електромагнітного випромінювання та аналізу електричних ланцюгів.

STFT ефективно вирішує проблеми визначення частотного спектра сигналів із частотними спектрами, що змінюються в часі. FT - це процес перетворення між часовою областю та частотою. Воно, за своєю суттю, є воротнім перетворенням, яке може перетворювати вихідні сигнали і перетворені сигнали один в одного. Характеристики FT не підходять для аналізу нестаціонарних сигналів, які можуть безпосередньо представляти мовленнєві сигнали. Мовленнєві сигнали змінюються повільно зі зміною часу, що можна розглядати як незмінні за короткий проміжок часу.

STFT - це короткочасний аналіз інформації частотної області, що широко використовується інструмент для обробки мовленнєвих сигналів. Є тільки одна різниця між FT та STFT. STFT полягає в тому, щоб розділити мовленнєві сигнали на досить маленькі фрагменти, щоб мовленнєві сигнали можна розглядати як стабільні. Формула STFT визначається як:

$$STFT(t,w) = \int_{-\infty}^{+\infty} s(x)\gamma(x-t)e^{-jwx}dx. \quad (2.5)$$

Тут *γ(x)* – вікно шириною *N* точок, *s(x)* позначає вхідний сигнал в момент часу *x*, який необхідно перетворити, а *STFT (t,* w) - двовимірна комплексна функція, яка представляє амплітуду і фазу, які змінюються за часом і частотою.

При придушенні акустичного еха, вилучення ознак - це незамінний процес, і вибір ознак вплине навчання моделі ехоподавления. З погляду одержаних базових одиниць, показники мовлення переважно діляться на показники рівня одиниці часу-частоти і показники рівня кадру. Характеристики рівня кадру витягуються із кадру сигналів.

Модель ехоподавления може краще відображати інформацію мовлення з допомогою функцій лише на рівні кадру і може аналізувати просторово-часову структуру мовлення у всіх напрямах під час навчання. Амплітуда спектру, отримана за допомогою STFT, може спростити навчання моделі та прискорити її збіжність.

2.5 Метрики якості моделі

Для оцінки якості моделі у роботі використовуються три метрики.

1. ERLE. Вимірює, наскільки добре акустичне ехо видаляється зі змішаного сигналу і визначається за формулою [23]:

$$ERLE = \lim_{n\to\infty} 10log_{10}\left(\frac{E[y^2(n)]}{E[e^2(n)]}\right) \quad (2.6)$$

де E – середнє значення. Чим вище цей показник, тим краще, він говорить про менше залишкове ехо.

2. PESQ. Загальноприйнята метрика для оцінки якості мовлення, яка порівнює вихідний сигнал *s(n)* з отриманою для нього оцінкою $\hat{s}(n)$. Процедура

обчислення PESQ є досить складною, її опис наводиться, зокрема, у роботі [24]. У більшості практичних випадків величина PESQ набуває значення в діапазоні від 0.5 до 4.5. Вищий бал вказує на кращу якість.

3. STOI. Практично завжди йде поруч із PESQ і відповідає за «розуміння» мовлення [25]. Зазвичай значення STOI знаходяться в діапазоні [0, 1] і, можна вважати, що це означає відсоток правильності роботи моделі.

2.6 Вхідні дані для моделі

Використовується набір даних TIMIT - один із відомих наборів даних для акустико-фонетичних досліджень, а також для розробки та оцінки систем автоматичного розпізнавання мовлення. TIMIT містить записи 630 носіїв 8 головних діалектів англійської (у США) мови, кожен із яких читає десять речень з фонетично багатим звучанням. Набір даних записаний із частотою 16 KHz у Texas Instruments, Inc, Массачусетським технологічним інститутом [26].

Вихідні дані є аудофайли з бази даних TIMIT. З цих даних (випадково) вибиралися сигнали ближнього кінця *s(n),* дальнього *x(n)* і формували відповідні сигнали мікрофона *y(n).* До цих сигналів, передискретизованих з частоти 16 кГц до 8 кГц (з метою зменшення часу обробки даних) було застосовано STFT з вікном Ханнінга шириною 256 пікселів, що відповідає тимчасовій довжині 32 мілісекунди та 129 елементам дозволу за частотою. Для збільшення навчальної вибірки було проведено аугментація даних, що полягає у перекритті тимчасових сигналів на 50%.

Отримані спектрограми Y сигналів мікрофона $y(n)$ були розбиті на блоки 100*129 для того, щоб нейронна мережа завжди мала вхід фіксованого розміру (додаток А) . Підсумковий обсяг навчальної вибірки становив 13606 зразків, обсяг тестових даних - 3093 зразка, обсяг валідаційного набору даних - 1855 зразків. Таким чином, набір даних розбитий на тренувальний, тестовий, валідаційний датасети у співвідношенні приблизно 75% - 15% - 10%.

Також були знайдені спектрограми мети $s(n)$ та перешкоди $d(n)$, які використовувалися, щоб знайти маску IBM, що є виходом моделі для входу Y.

2.7 Вихід моделі

Виходом нейронної мережі (метою навчання, target) є IBM - одна з масок, що найчастіше використовуються в задачах розпізнавання мовлення. IBM визначається як [26]:

$$IBM(t,f) = \begin{cases} 1 & if \ \frac{S_T(t,f)}{S_I(t,f)} > 1 \\ 0 & otherwise \end{cases}. \quad (2.7)$$

Тут $S_T = S_T(t, f)$ та $S_I = S_I(t, f)$ - спектрограми мети $s(n)$ і перешкоди $d(n)$, відповідно. Деякі дослідники, наприклад [27], вважають, що з використанні IBM відновлена мова звучить не природно, проте розбірливість вимови у ній дуже хороша.

Якщо ми окреслимо Y спектрограму сигналу мікрофона, то

$$Y = S_T + S_I, \quad (2.8)$$

і, за допомогою маски IBM (додаток А) можна відновити спектрограму корисного сигналу $s(n)$ [26]:

$$S_T = IBM \odot Y \quad (2.9)$$

Тут оператор $\odot$ є поелементним множенням.

За відомою спектрограмою для $s(n)$ можна відновити сам сигнал через застосування зворотного перетворення Фур'є.

РОЗДІЛ 3. ПРАКТИЧНА ЧАСТИНА

3.1 Опис моделі BLSTM+clustering

Модель BLSTM має два шари. Кожен шар містить дві LSTM з 300 нейронами, одна з яких обробляє сигнал у прямому напрямку, а інша - у зворотному. Вихідний повнозв'язний шар має сигмоїдну функцію активації, і діапазон значень [0, 1], який легко (з установкою порогового значення 0.5) трансформується в дискретний вихід з нулів і одиниць, відповідний IBM масці.

Для навчання мережі був обраний оптимізатор adam, як функція втрат використовувалася середньоквадратична помилка MSE (Mean Square Error). Швидкість навчання дорівнювала 0.01. Кількість епох навчання дорівнювала 100 (додаток Б) .

Результати описаної моделі чистої BLSTM виявилися незадовільними (вони, як і результати для інших моделей, представлені в п. 3.4), тому було вирішено використовувати додатково глибоку кластеризацію.

На рис. 3.1 наведено структуру нейронної мережі BLSTM+clustering. У цьому випадку ми збільшили розміри матриці ваг і зміщення в три рази на останньому шарі нейронної мережі і тепер її вихід є матрицею розміру (12900, 3) (на відміну від матриці (12900, 1) для моделі BLSTM). Далі застосовується алгоритм кластеризації до 12900 точок у тривимірному просторі. Розділяючи дані на два класи, отримуємо вектор з нулів та одиниць, який потім перетворимо на матрицю, що відповідає масці IBM. Як алгоритми кластеризації у роботі використовувалися відомі алгоритми K-Means, Mean-shift, EM.

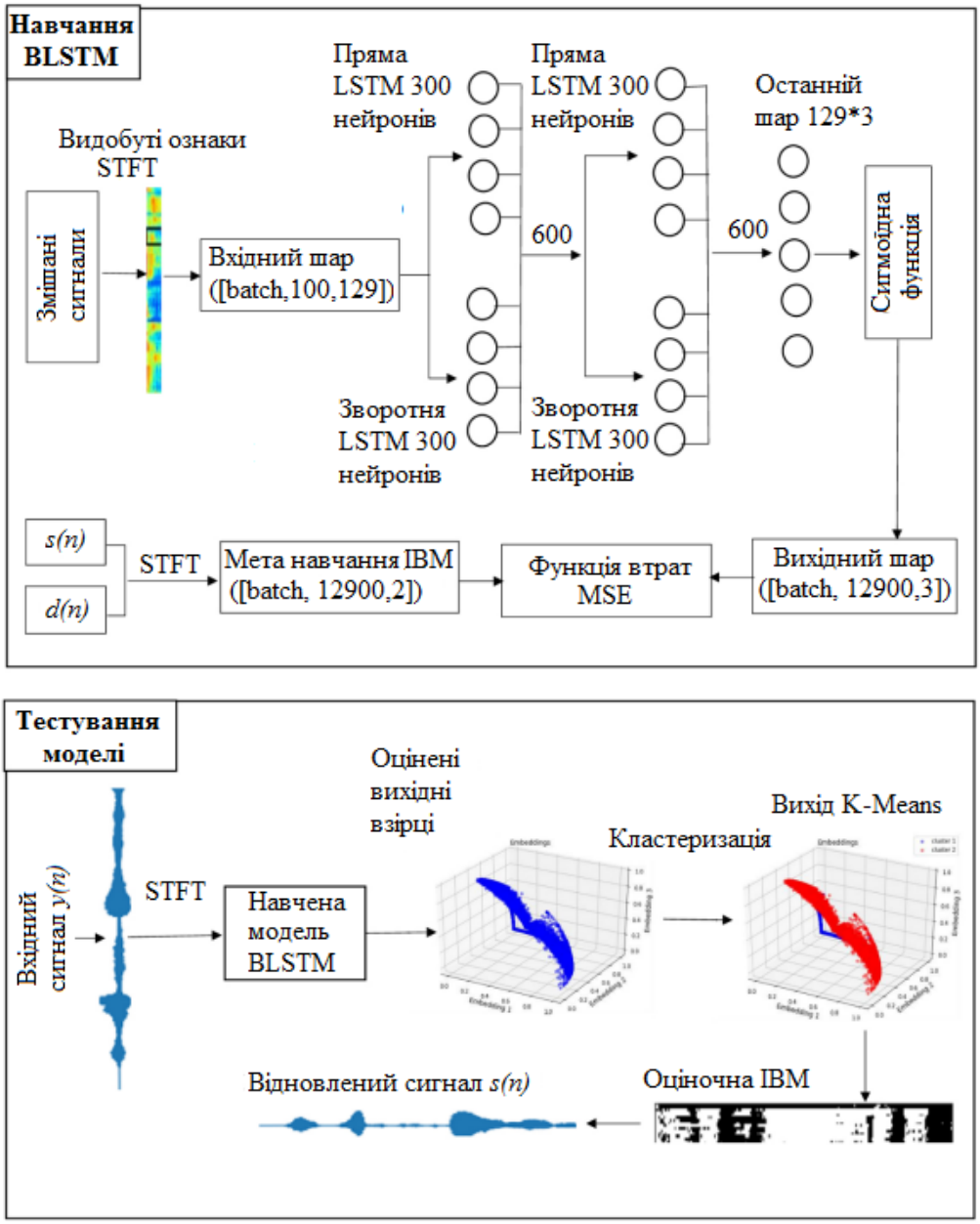

Рисунок 3.1 – Структура пропонованої моделі

3.2 Моделювання exo-сигналу в приміщенні

Змоделюємо кімнату, до якої під'єднано декілька джерел звуку та мікрофони. Метод моделювання характеристики RIR між джерелами та мікрофонами за умовчанням є методом джерела зображення ISM [28], який розглядає стіни як ідеальні відбивачі. Крім цього, доступний експериментальний гібридний симулятор, заснований на трасуванні променів RT (Ray Tracing) [29, 30] . RT краще вловлює пізніші відображення, а також може моделювати такі ефекти, як розсіювання.

Далі сигнали мікрофона створюються шляхом згортки аудіосемплів, пов'язаних із джерелами з відповідним RIR.

3.2.1 Моделювання приміщення shoebox за допомогою методу ISM

Ми використовуємо ISM, щоб знайти всі джерела зображень до максимального вказаного порядку, і RIRs генеруються з їхньої позиції.

Shoebox rooms - паралелепіпедні приміщення з 4 або 6 стінами (у 2D та 3D відповідно), все під прямим кутом. Вони визначаються одним вектором, який містить довжину стіни. Їхня перевага в тому, що їх легко визначати та дуже ефективно моделювати.

Використовуємо формулу Себіна [31] для визначення поглинання енергії стінкою та максимального порядку ISM, необхідного для досягнення бажаного часу реверберації (rt60, тобто час необхідне зменшення RIR на 60 dB).

Метод pra.ShoeBox (лістинг 3.1) має 4 аргументи, перший аргумент – розмір кімнати, визначаємо кімнату розміром (9, 7.5, 3.5) m. Другий аргумент - частота вибірки fs, з якою створюватиметься RIR. Зауважте, що значення fs за замовчуванням становить 8 кГц. Третій аргумент - матеріал стіни, який сам приймає поглинання як параметр. Четвертий та останній аргумент – максимальна кількість відображень, дозволених у ISM.

Лістинг 3.1 - Метод pra.ShoeBox

```
import pyroomacoustics as pra
rt60 = 0.5
room_dim = [9, 7.5, 3.5]
e_absorption, max_order = pra.inverse_sabine(rt60, room_dim)
room = pra.ShoeBox(room_dim, fs=8000, materials =
pra.Material(e_absorption ), max_order=max_order)
```

3.2.2 Додавання джерел та мікрофонів

Лістинг 3.2 ілюструє створення джерел. Джерела приймають своє місце розміщення як єдиний обов'язковий аргумент, а сигнал і час початку - як необов'язкові аргументи. Створимо джерело, розташоване в (2.5, 3.73, 1.76) m всередині кімнати, яке озвучуватиме вміст wav- файлу (взяте з бази даних

TIMIT), починаючи з 1.3 секунди в процесі моделювання (лістинг 3.2). Аргумент ключового слова signal для методу add_source () повинен бути одновимірним масивом numpy.ndarray, що містить бажаний звуковий сигнал.

Розташування мікрофонів у масиві має бути зазначено в numpy nd-array розміру (ndim, nmics), тобто кожен стовпець містить координати одного мікрофона. Створюємо масив з одним мікрофоном, розміщеним у (6.3, 4.87, 1.2) m усередині кімнати.

Лістинг 3.2 - Додавання джерел та мікрофонів

```
from scipy.io import wavfile
import numpy as np
_, audio = wavfile.read('/ TIMIT/SA1.WAV.wav')
room.add_source([2.5, 3.73, 1.76], signal=audio, delay=1.3)
mic_locs = np.c_[
[6.3, 4.87, 1.2],
]
room.add_microphone_array(mic_locs)
```

3.2.3 Створення імпульсної характеристики приміщення

У цьому етапі RIRs просто створюються шляхом виклику ISM через compute_rir() (лістинг 3.3) . Ця функція згенерує всі джерела зображень у потрібному порядку і буде використовувати їх для створення RIRs, які будуть зберігатися в атрибуті room.rir, що є списком списків, так що зовнішній список відноситься до мікрофонів, а внутрішній список до джерел.

Лістинг 3.3 - Створення RIR

```
room.compute rir()
import matplotlib.pyplot as plt
plt.plot(room.rir[0][0])
plt.xlabel("Number of samples")
plt.ylabel("Amplitude")
plt.show()
```

На рис. 3.2 показаний RIR, отриманий за допомогою методу ISM.

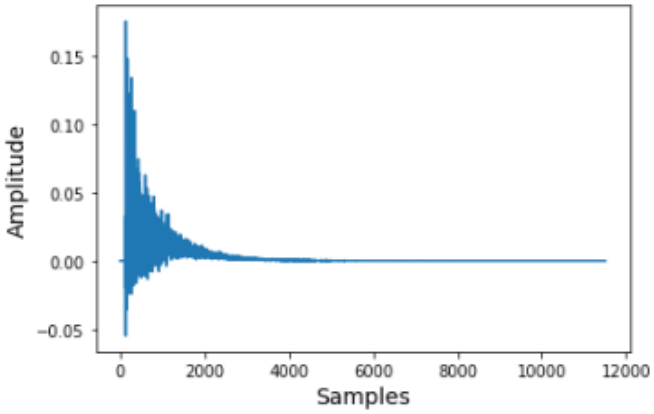

Рисунок 3.2 – RIR, отриманий методом ISM

3.2.4 Формування exo-сигналу

При виклику simulate (), згортка сигналу кожного джерела (якщо не None) буде виконуватися з відповідним RIR (лістинг 3.4) . Вихідні дані згорток сумуватимуться на мікрофонах. Результат зберігається в атрибуті room.mic_array.signals, де кожен рядок відповідає одному мікрофону.

Лістинг 4 - Формування exo-сигналу

```
room.simulate()
plt.plot(room.mic array.signals[0,:])
plt.xlabel("Number of samples")
```

На рис. 3.3 наведено приклад сигналу джерела (Source signal), а також exo-сигналу (Echo signal) результатом моделювання за допомогою RIR.

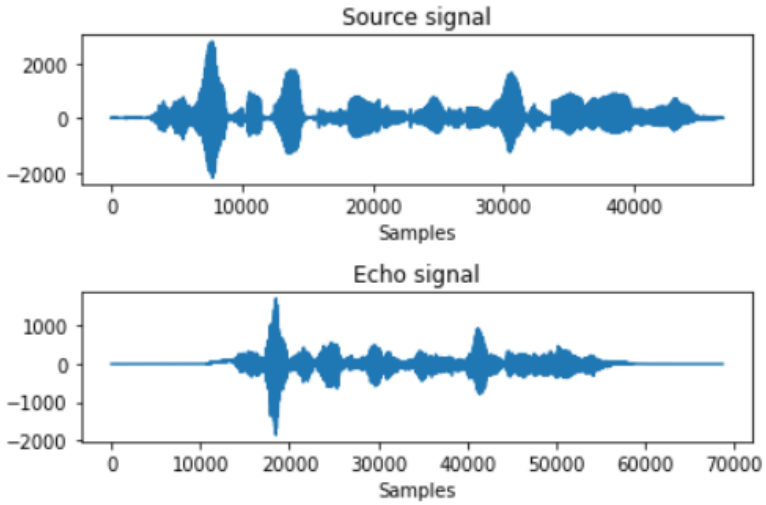

Рисунок 3.3 – Сигнал джерела та exo-сигнал, отриманий за допомогою RIR

Це важливий етап у нашій роботі, де ми повинні сформувати ехо-сигнали, які потім будуть об'єднані із ближнім сигналом.

3.3 Технології для реалізації моделі

Як технологію для опису та реалізації моделі обрано бібліотеку для задач машинного навчання Tensorflow [32], розроблена фахівцями науково-дослідного підрозділу Google Brain. Є відкритою для чисельних обчислень з використанням графів потоків даних. Tensorflow має високорівневий API для моделювання складних архітектур нейромереж, таких як рекурентні або згорткові.

Для реалізації моделі машинного навчання необхідно описати граф, вершини у якому є математичні операції, а ребра у графі є тензори - масиви довільної розмірності, де тип базового елемента заданий чи виведений на момент побудови графа. Робота з фреймворком Tensorflow ділиться на два основні етапи: конструювання графа операцій та інтерпретація графа конфігурованої сесією Tensorflow, яка специфікує, наприклад, параметри обчислень на графічних прискорювачах та конфігурацію логування.

Робота виконана з використанням середовища Python у хмарі Google Colab.

3.4 Оцінка ефективності моделей з двома розмовами

3.4.1 Оцінка ефективності моделі BLSTM

Для навчання, валідації та тестування з вихідного набору даних з 630 носіями були випадково обрані записи 462, 68 і 100 осіб, відповідно. Оскільки кожна людина читає 10 пропозицій, ми маємо 4620 аудіофайлів для навчання, 680 для валідації та 1000 для тестування. Випадково обрана пара з цього набору є, як правило, аудіофайлами з мовленням різних людей, які ми беремо як сигнали ближнього $s(n)$ та далекого кінця $x(n)$. З сигналу $x(n)$ шляхом згортки з імпульсною характеристикою $h(n)$ приміщення RIR формується ехо-сигнал $d(n)$. Змішуючи $d(n)$ із $s(n)$, отримуємо сигнал мікрофона. Таким чином, у нас є 2310,

340 та 500 пар аудіофайлів для навчання, валідації та тестування. Оскільки тривалість аудіофайлів різна, то з кожного з них ми отримуємо різну кількість спектрограм (у середньому близько 3, але з урахуванням аугментації і перекриття в 50%, близько 6). Остаточно, обсяг навчальної, валідаційної та тестової вибірок склав у наших експериментах 13606, 1855 та 3093, відповідно.

RIR генерується за часу реверберації T60 = 0.5c (час, необхідний для зменшення RIR на 60 dB) з використанням методу ISM. Розмір кімнати для моделювання становить (9, 7.5, 3.5) m, мікрофон знаходиться в позиції (6.3, 4.87, 1.2) m всередині кімнати, джерело перешкоди (2.5, 3.73, 1.76) m. Джерело перешкоди озвучує вміст wav -файлу (з TIMIT), починаючи з 1.3 секунди.

SER обчислюється за формулою (3.1):

$$SER = 10log_{10}\left(\frac{E[s^2(n)]}{E[d^2(n)]}\right). \quad (3.1)$$

Далі в роботі буде наведено метрики оцінки якості моделей для випадків SER=6 дБ.

На рис. 3.4 представлені графіки зміни метрики Accuracy (частки правильних відповідей) для тренувальної та валідаційної вибірок у процесі 100 епох навчання.

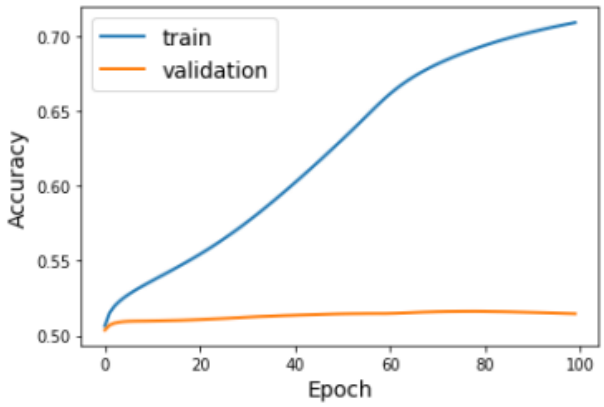

Рисунок 3.4 – Оцінка ефективності моделі BLSTM

Видно, що метрика досягла майже постійного порогу, який, однак, становить трохи більше 50% для валідаційної вибірки, що зовсім незадовільно.

На рис. 3.5 показано приклад спектрограми сигналу мікрофона, оціночної маски IBM, що є цільовим виходом моделі BLSTM, і справжньої маски (Real IBM).

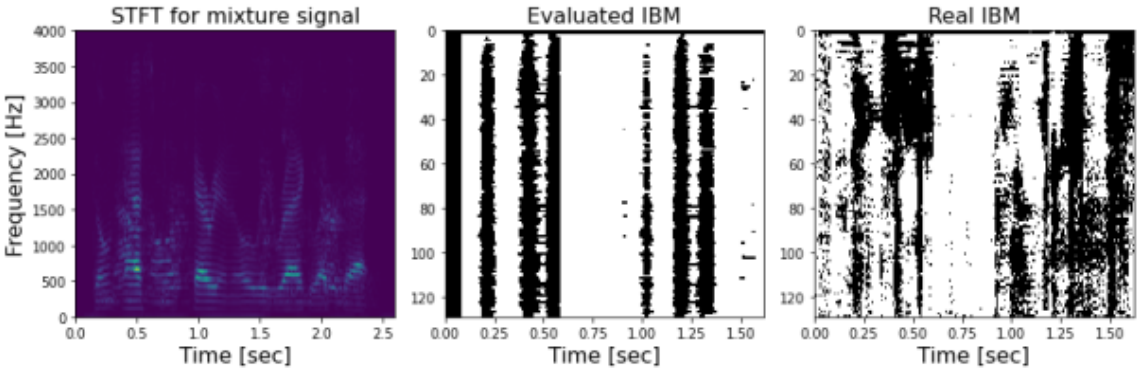

Рисунок 3.5 – Спектрограма сигналу мікрофона, оціночна маска IBM, яка отримана методом BLSTM та справжня маска IBM

На рис. 3.6 показані сигнал входу BLSTM *y(n)* (Mixture signal), сигнал ближнього кінця *s(n)* (Near-end signal), ресинтезований цільовий сигнал $\hat{s}(n)$ (Target signal), і exo *d(n)* (Echo).

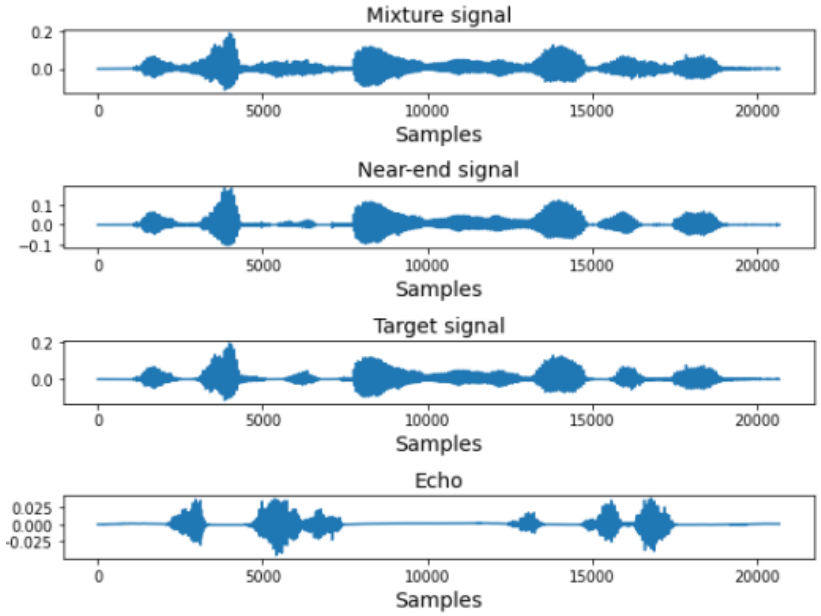

Рисунок 3.6 – Вхідний та вихідний сигнали моделі BLSTM

Можна бачити, що сигнали *s(n)* та $\hat{s}(n)$ дуже схожі. Значення метрик PESQ та STOI склало 1.03 та 0.846 відповідно.

3.4.2 Оцінка ефективності моделі BLSTM+k-Means

Перейдемо до побудови запропонованої моделі після додавання кластеризації на виході BLSTM. На рис. 3.7 представлені графіки зміни метрики Accuracy для тренувальної та валідаційної вибірок у процесі 100 епох навчання. Зазначимо, що на відміну від аналогічного рис. 3.4 для простої моделі BLSTM, досягнуто точність близько 78% на валідаційній вибірці.

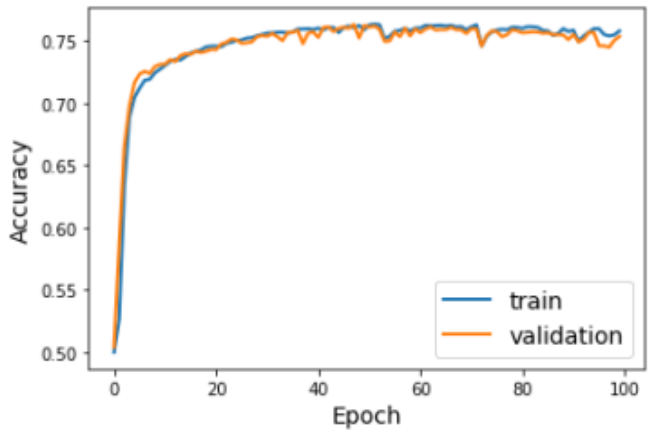

Рисунок 3.7 – Оцінка ефективності моделі BLSTM+k-Means

На рис. 3.8 показані точки в тривимірному просторі, що є виходом навчальної моделі, а також показано розбиття цих точок на два класи, отримане за допомогою алгоритму K- Means

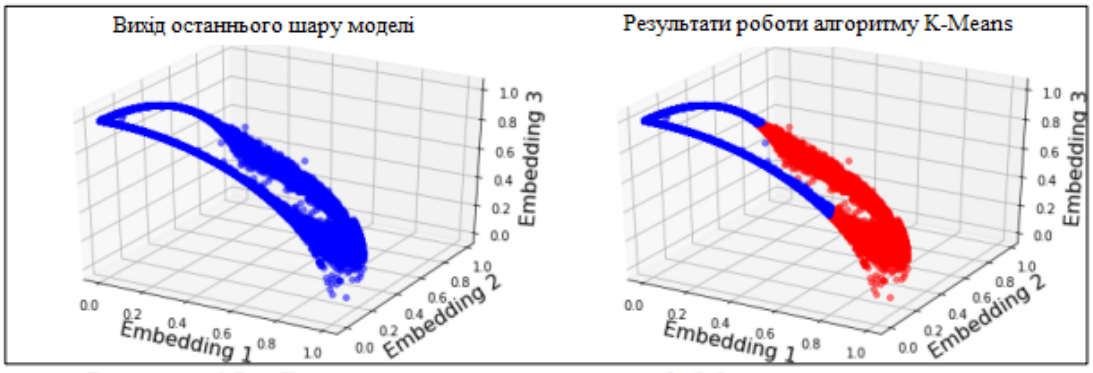

Рисунок 3.8 – Вихід моделі та результат k-Means кластеризації

На рис. 3.9 наведено приклад спектрограми сигналу мікрофона, оціночної маски 1BM, що є цільовим виходом BLSTM+k-Means, а також істинної маски ((Real IBM).

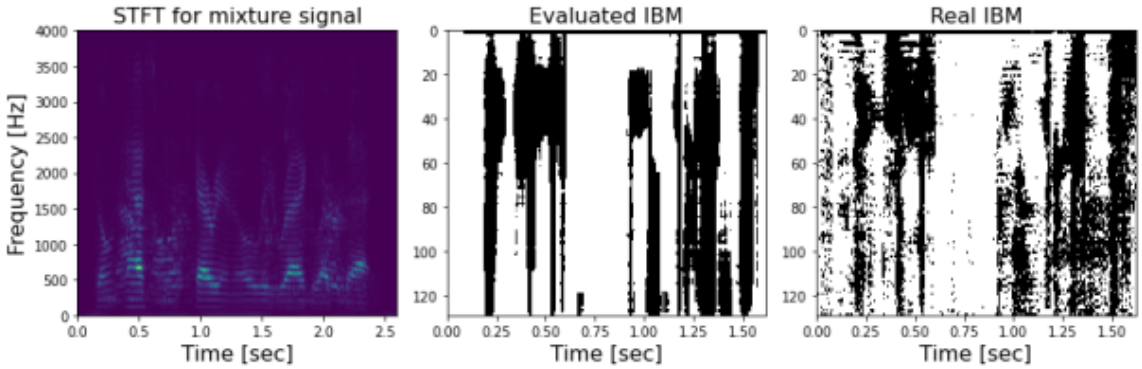

Рисунок 3.9 – Спектрограма сигналу мікрофона, оціночна маска IBM, яка отримана методом BLSTM+k-Means та справжня маска IBM

На рис. 3.10 показані сигнал входу BLSTM+k-Means, сигнал ближнього кінця, цільовий ресинтезований сигнал і exo. Значення метрик PESQ і STOI, що характеризують якість відновлення сигналу, склало 2.1 і 0.911, відповідно, і перевищило їх значення у випадку простої моделі BLSTM.

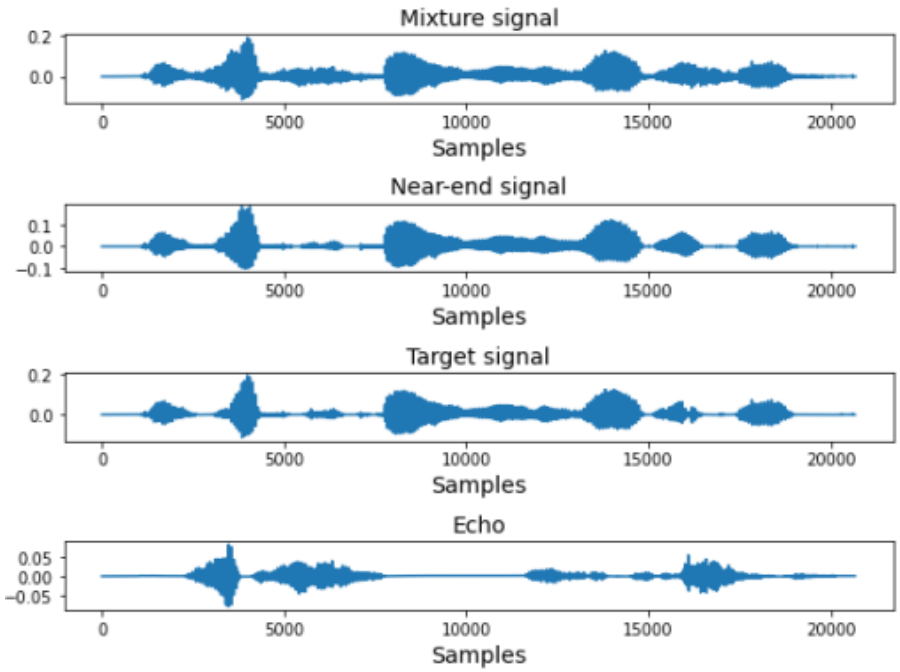

Рисунок 3.10 – Вхідний та вихідний сигнали моделі BLSTM+k-Means

3.4.3 Оцінка ефективності моделі BLSTM+EM

На рис. 3.11 показані точки у тривимірному просторі, що є виходом навчальної моделі, а також показано розбиття цих точок на два класи, отримане за допомогою алгоритму EM.

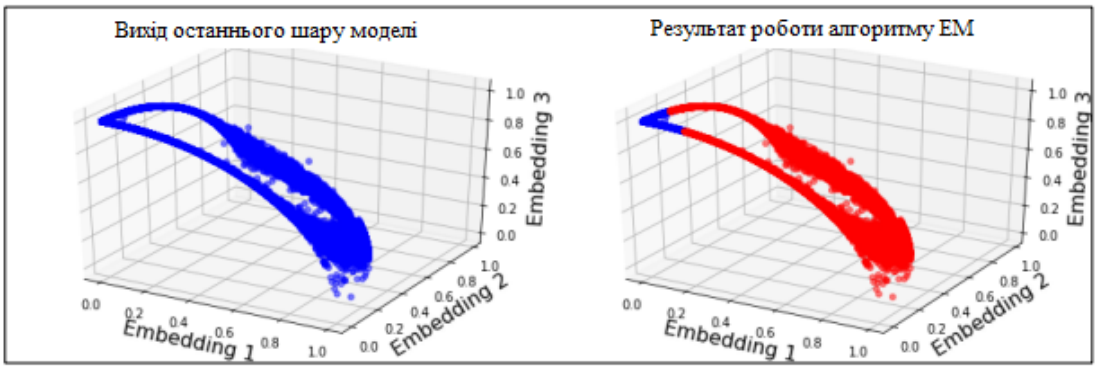

Рисунок 3.11 – Вихід моделі та результат EM кластеризації

На рис. 3.12 приведено приклад спектрограми сигналу мікрофона, оціночної маски IBM, яка є цільовим виходом моделі BLSTM+EM, а також справжньої маски (Real IBM).

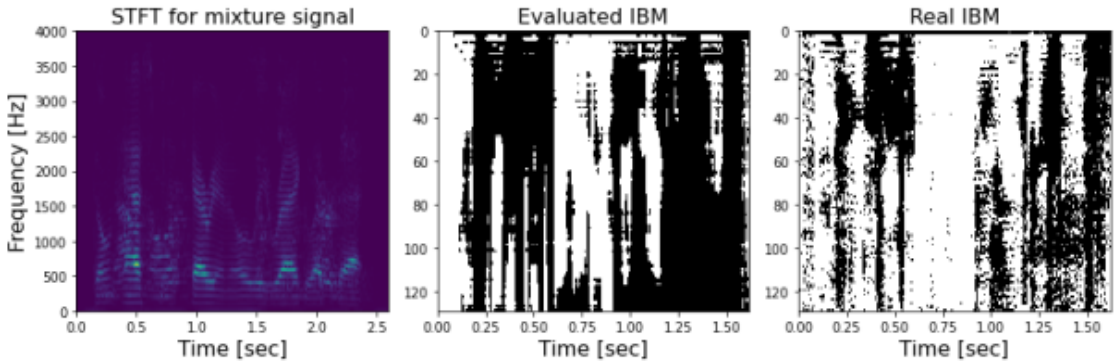

Рисунок 3.12 – Спектрограма сигналу мікрофона, оціночна маска IBM, яка отримана методом BLSTM+EM та істинна маска IBM

На рис. 3.13 показані сигнал входу BLSTM+EM, сигнал ближнього кінця, ресинтезований цільовий сигнал та exo. Значення метрик PESQ і STOI склало 0.91 і 0.714, відповідно, і не перевищило їх значення у разі простої моделі BLSTM.

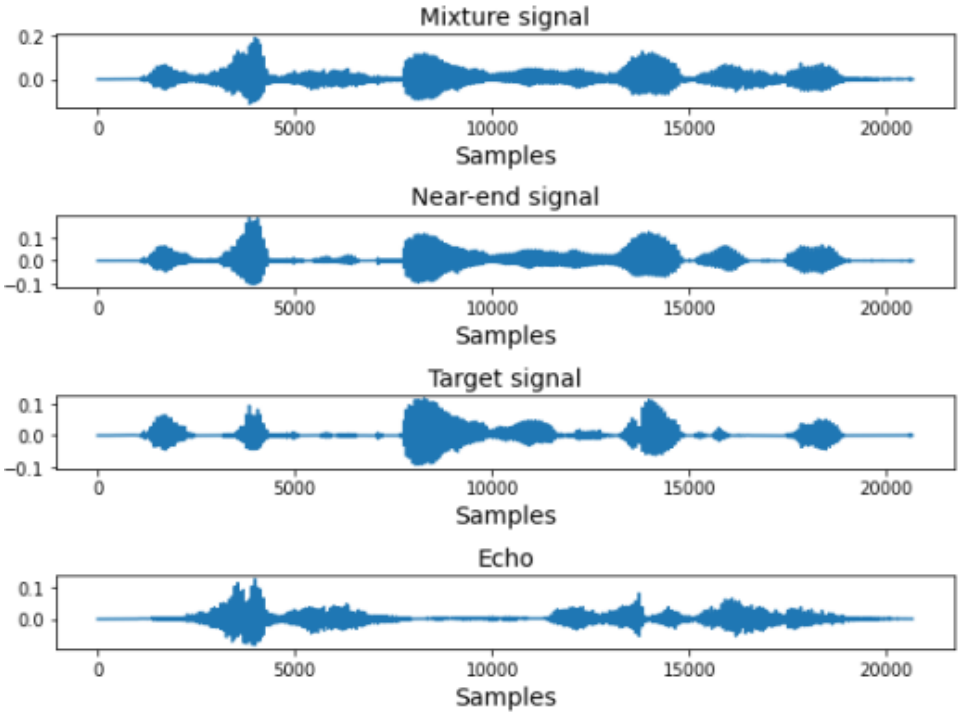

Рисунок 3.13 – Вхідний та вихідний сигнали моделі BLSTM+EM

3.4.4 Оцінка ефективності моделі BLSTM+Mean-Shift

На рис. 3.14 показані точки у тривимірному просторі, що є виходом навчальної моделі, а також показано розбиття цих точок на два класи, отримане за допомогою алгоритму Mean-Shift.

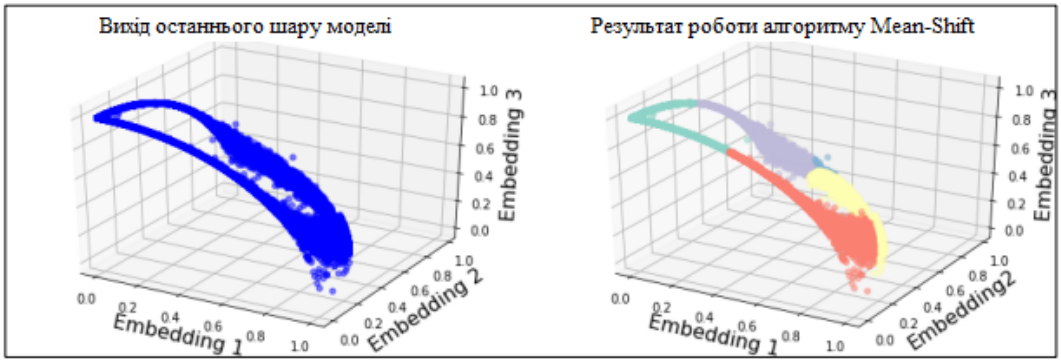

Рисунок 3.14 – Вихід моделі та результат Mean-Shift кластеризації

На рис. 3.15 показаний приклад спектрограми сигналу мікрофона, оціночної маски IBM, що є цільовим виходом моделі BLSTM+Mean-Shift, і справжньої маски (Real IBM).

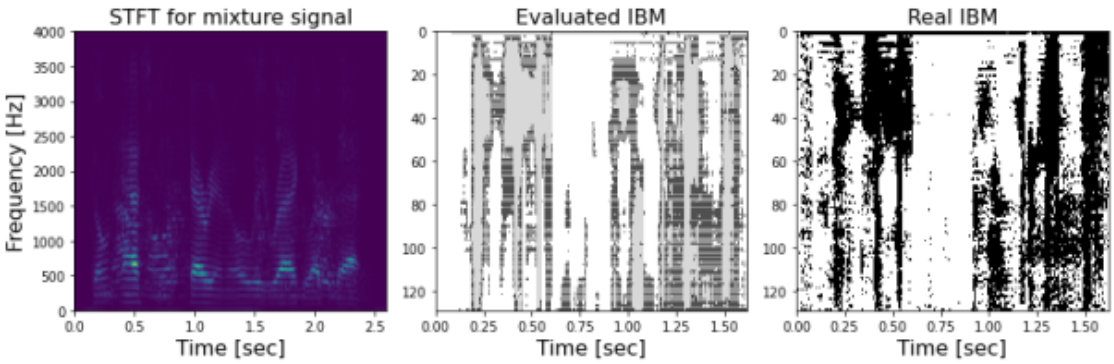

Рисунок 3.15 – Спектрограма сигналу мікрофона, оціночна маска IBM, яка отримана методом BLSTM+Mean-Shift та істинна маска IBM

На рис. 3.16 показані сигнал входу BLSTM+Mean-Shift, сигнал ближнього кінця, ресинтезований цільовий сигнал та exo. Значення метрик PESQ та STOI, склало 1.17 та 0.808, відповідно, і PESQ перевищило її значення у випадку простої моделі BLSTM.

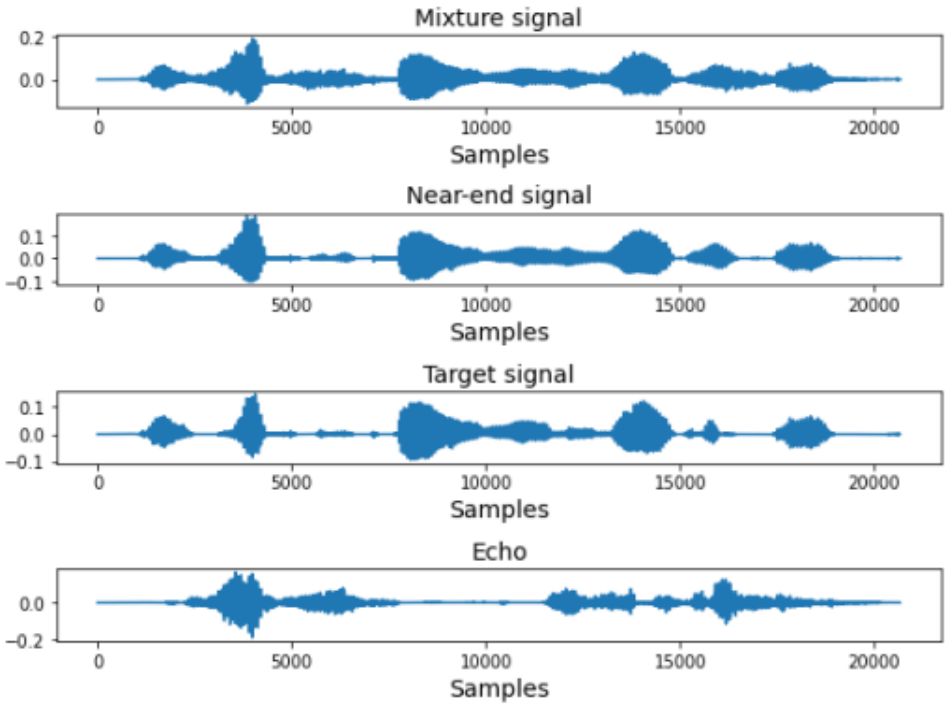

Рисунок 3.16 – Вхідний та вихідний сигнали моделі BLSTM+ Mean-Shift

3.4.5 Порівняння ефективності моделей

У табл. 3.1 наведені значення метрик ERLE, PESQ і STOI, отриманих чотирма методами, що розглядаються при SER=6 дБ.

Таблиця 3.1 - Порівняння моделей при SER = 6 дБ

Метод	ERLE	PESQ	STOI
BLSTM	6.8	1.03	0.846
BLSTM+EM	3.5	0.91	0.714
BLSTM+Mean-Shift	-2.6	1.17	0.808
BLSTM+K-Means	**8.1**	**2.1**	**0.911**

Зазначимо, що для обчислення метрик використовувалися дані, які брали участь у процесі навчання. Як бачимо, використання алгоритму k-Means поліпшило всі показники, тоді як інші алгоритми кластеризації майже завжди погіршують роботи моделі BLSTM.

Табл. 3.2 містить метрики ефективності моделей для випадку SER =10 дБ.

Таблиця 3. 2 - Порівняння моделей при SER = 10 дБ

Метод	ERLE	PESQ	STOI
BLSTM	8.7	2.23	0.865
BLSTM+EM	5.3	1.59	0.770
BLSTM+Mean-Shift	-1.8	2.14	0.846
BLS TM+K-Means	**11.2**	**2.65**	**0.924**

Видно, що і в цьому випадку додавання k-Means до BLSTM показує покращення значень ERLE десь на 2.5 дБ, PESQ на 0.42 та STOI на 0.059. Таким чином, метрика STOI, що характеризує розбірливість мовлення, покращала на 7%, а метрика PESQ, що характеризує якість відновлення мови, на 18.8%. Використання Mean-Shift та EM не покращило продуктивність моделі BLSTM.

3.5 Оцінка ефективності моделей з однією розмовою

3.5.1 Оцінка ефективності моделі BLSTM

Припустимо, що сигнал далекого й ближнього кінця однаковий, тобто у

нас ситуація з однією розмовою. Наше завдання полягає в тому, щоб видалити exo з ближнього кінця, яке моделюється так само, якби у нас було дві розмови.

У цьому випадку сигнал мікрофона визначається таким чином:

$$y(t) = s(t) + s(t) * h(t) = s(t) + \int s(t)h(t-\tau)d\tau \qquad (3.2)$$

Тут $s(t)$ представляє сигнал ближнього кінця. Сигнал далекого кінця $x(t)$ дорівнює $s(t)$; $s(t) * h(t)$ є exo-сигнал $d(t)$.

Випадково обраний аудіофайл з бази даних TIMIT є сигналом ближнього кінця $s(n)$. Далі із сигналу $s(n)$ шляхом згортки з $h(n)$ формується exo- сигнал $d(n)$. Розмір кімнати для моделювання становить (9, 7.5, 3.5) m, мікрофон знаходиться в позиції (6.3, 4.87, 1.2) m всередині кімнати, джерело перешкоди (2.5, 3.73, 1.76) m. Джерело перешкоди озвучує вміст wav -файлу (з TIMIT), починаючи з 1.3 секунди.

Далі в роботі буде наведено метрики оцінки якості моделей для випадків SER=5 дБ.

На рис. 3.17 показаний приклад спектрограми сигналу мікрофона, оцінної маски IBM, що є цільовим виходом моделі BLSTM, і справжньої маски (Real IBM).

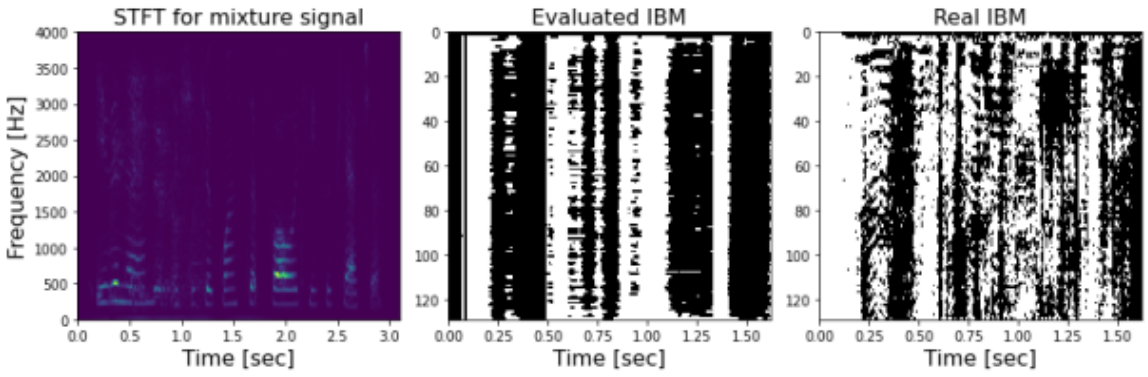

Рисунок 3.17 – Спектрограма сигналу мікрофона, оціночна маска IBM, яка отримана методом BLSTM та істинна маска IBM

На рис. 3.18 показані сигнал входу BLSTM, сигнал ближнього кінця $s(n)$

(Near-end signal), ресинтезований цільовий сигнал $\hat{s}(n)$ (Target signal), і exo *d(n)* (Echo). Можна бачити, що значення метрик PESQ та STOI склало 0.98 та 0.759, відповідно.

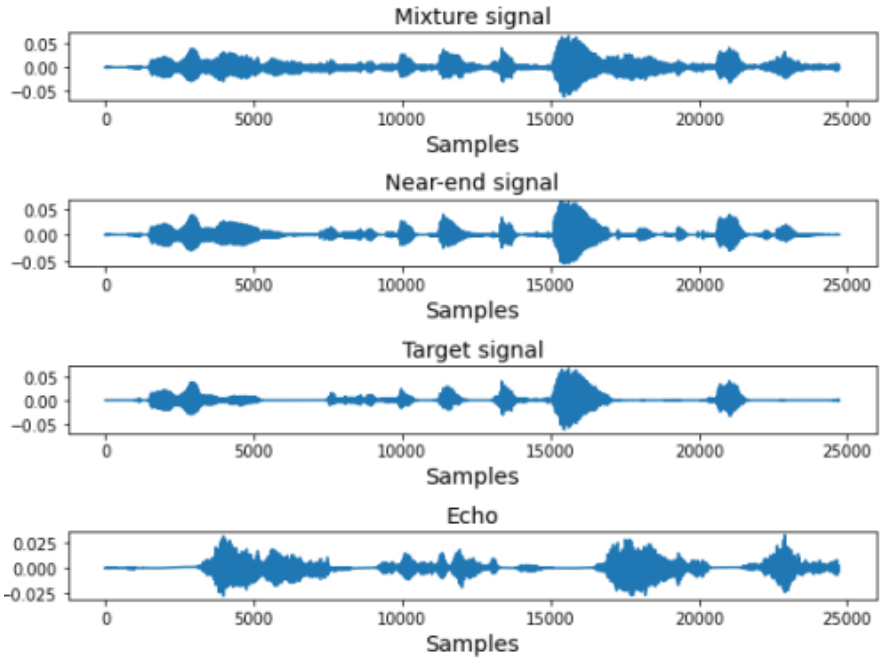

Рисунок 3.18 – Вхідний та вихідний сигнали моделі BLSTM

3.5.2 Оцінка ефективності моделі BLSTM +K-Means

На рис. 3.19 показані точки у тривимірному просторі, що є виходом навчальної моделі, а також показано розбиття цих точок на два класи, отримане при допомоз алгоритму K-Means.

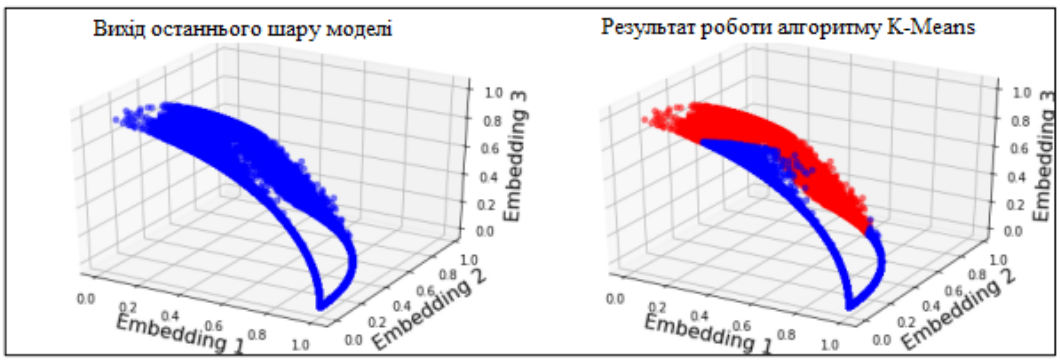

Рисунок 3.19 – Вихід моделі та результат K-Means кластеризації

На рис. 3.20 показаний приклад спектрограми сигналу мікрофона,

оціночної маски IBM, котра є цільовим виходом BLSTM+K-Means, а також справжньої маски (Real IBM).

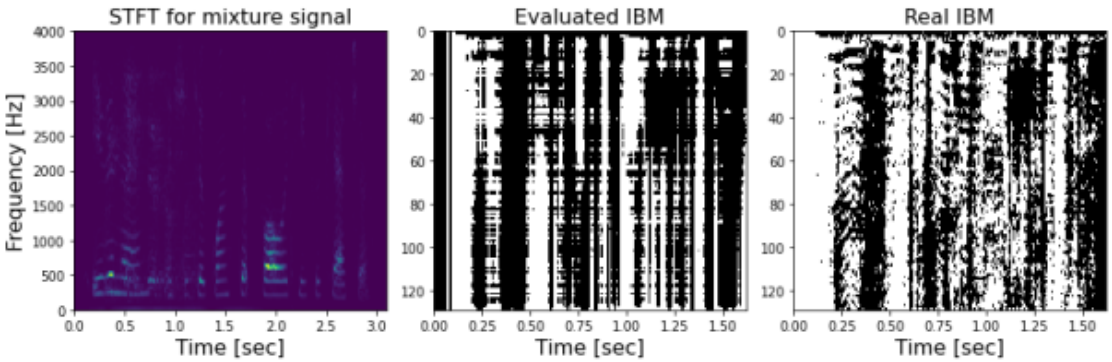

Рисунок 3.20 – Спектрограма сигналу мікрофона, оціночна маска IBM, яка отримана методом BLSTM+K-Means та істинна маска IBM

На рис. 3.21 показані сигнал входу BLSTM+K-Means, сигнал ближнього кінця, цільовий ресинтезований сигнал і exo. Значення метрик PESQ і STOI склало 1.91 і 0.894, відповідно, і перевищило їх значення у випадку простої моделі BLSTM.

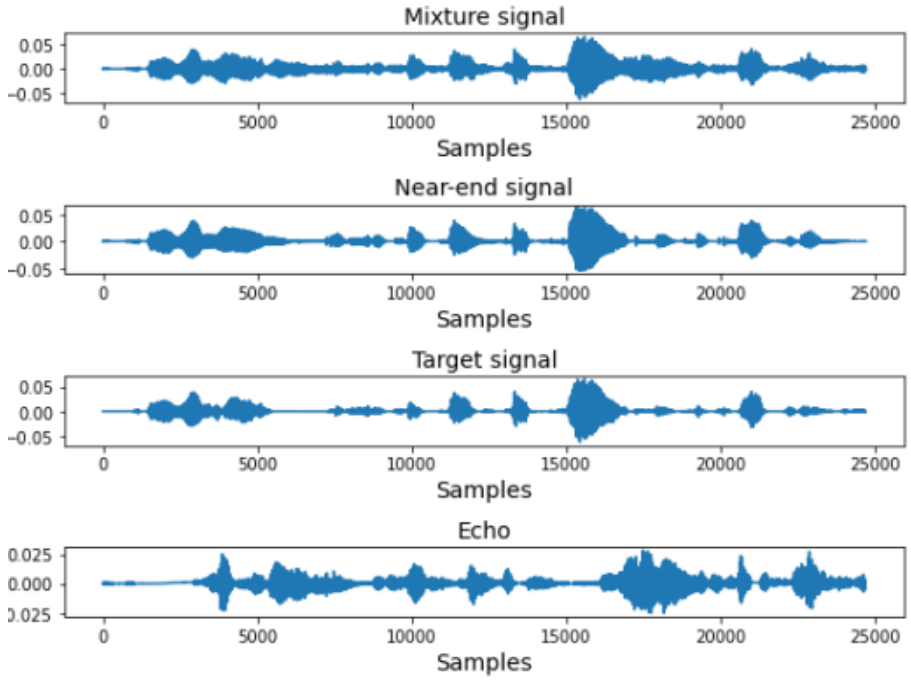

Рисунок 3.21 – Вхідний та вихідний сигнали моделі BLSTM+K-Means

Аналогічно було проаналізовано алгоритми BLSTM+EM,

BLSTM+MeanShift, засновані на використанні методів кластеризації EM та MeanShift. Отримані значення метрик наведено у табл. 3.3.

3.5.3 Порівняння ефективності моделей

У табл.3.3 показано метрики ефективності моделей для випадку SER = 5 дБ.

Таблиця 3.3 - Порівняння моделей при SER = 5 дБ

Метод	ERLE	PESQ	STOI
BLSTM	5.2	0.98	0.759
BLSTM+EM	7.17	1.76	0.873
BLSTM+Mean-Shift	-1.1	1.26	0.823
BLSTM+K-Means	**8.2**	**1.91**	**0.894**

Можна бачити, що і в цьому випадку K-Means, як і раніше, перевершує методи EM та Mean-Shift. Додавання k-Means до BLSTM показує покращення значень ERLE приблизно на 3 дБ, PESQ на 0.93 та STOI на 0.135. Таким чином, метрика STOI 17,8%, а метрика PESQ на 94,8%.

Також зазначимо, що додавання EM до BLSTM показує покращення значень ERLE приблизно на 1.97 дБ, PESQ на 0.78 та STOI на 0.114. Таким чином, метрика STOI на 15%, а метрика PESQ – на 79.2% також покращилися при використанні методу Mean-Shift, за винятком метрики ERLE, яка стала гіршою.

РОЗДІЛ 4. БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ОХОРОНИ ПРАЦІ

4.1 Класифікація шкідливих та небезпечних виробничих факторів

Шкідливий виробничий фактор – небажане явище, яке супроводжує виробничий процес і вплив якого на працюючого може призвести до погіршення самопочуття, зниження працездатності, захворювання, виробничо зумовленого чи професійного, і навіть смерті, як результату захворювання. Небезпечний виробничий фактор – небажане явище, яке супроводжує виробничий процес і дія якого за певних умов може призвести до травми або іншого раптового погіршення здоров'я працівника (гострого отруєння, гострого захворювання) і навіть до раптової смерті [33].

Поділ несприятливих чинників виробничого середовища на шкідливі та небезпечні зумовлене різним характером їх дії на людський організм, тим, що вони потребують різних заходів та засобів для боротьби з ними та профілактики викликаних ними ушкоджень, а також рядом причин організаційного характеру. В той же час між шкідливими та небезпечними виробничими факторами інколи важко провести чітку межу. Один і той же чинник може викликати травму і профзахворювання (наприклад, високий рівень іонізуючого або теплового випромінювання може викликати опік або навіть призвести до миттєвої смерті, а довготривала дія порівняно невисокого рівня цих же факторів – до хвороби; пилинка, що потрапила в око, спричиняє травму, а пил, що осідає в легенях, – захворювання, що зветься пневмоконіоз). Через це всі несприятливі виробничі чинники часто розглядаються як єдине поняття – небезпечний та шкідливий виробничий фактор (НШВФ) [33]. За своїм походженням та природою дії всі НШВФ можна поділити на 5 груп: фізичні, хімічні, біологічні, психофізіологічні та соціальні До фізичних НШВФ відносяться машини та механізми або їх елементи, а також вироби, матеріали, заготовки тощо, які рухаються або обертаються; конструкції, які руйнуються; системи, устаткування або елементи обладнання, які знаходяться під підвищеним тиском; підвищена запиленість та

загазованість повітря; підвищена або понижена температура повітря, поверхонь приміщення, обладнання, матеріалів; підвищені рівні шуму, вібрації, ультразвуку, інфразвуку; підвищений або понижений барометричний тиск та його різкі коливання; підвищена та понижена вологість; підвищена швидкість руху та підвищена іонізація повітря; підвищений рівень іонізуючих випромінювань; підвищене значення напруги в електричній мережі; підвищені рівні статичної електрики, електромагнітних випромінювань; підвищена напруженість електричного, магнітного полів; відсутність або нестача світла; недостатня освітленість робочої зони; підвищена яскравість світла; понижена контрастність; прямий та віддзеркалений блиск; підвищена пульсація світлового потоку; підвищені рівні ультрафіолетової та інфрачервоної радіації; гострі крайки, зачипки, шершавість на поверхні заготовок, інструментів та обладнання; розташування робочого місця на значній висоті відносно землі (підлоги); слизька підлога; невагомість.

Хімічні НШВФ:

- за характером дії на організм людини поділяються на токсичні, задушливі, наркотичні, подразнюючі, сенсибілізуючі, канцерогенні, мутагенні та такі, що впливають на репродуктивну функцію;

- за шляхами проникнення в організм людини поділяються на такі, що потрапляють через: 1) органи дихання; 2) шлунково-кишковий тракт; 3) шкіряні покриви та слизова оболонка;

- які перебувають у різному агрегатному стані: 1)твердому 2)газоподібному 3)рідкому.

Біологічні НШВФ – це: - патогенні мікроорганізми (бактерії, віруси, рикетсії, спірохети, грибки, найпростіші) та продукти їхньої життєдіяльності; - макроорганізми (тварини та рослини) та продукти їхньої життєдіяльності. До психофізіологічних НШВФ відносяться фізичні (статичні та динамічні) перевантаження і нервово-психічні перевантаження (розумове перенапруження, перенапруження аналізаторів, монотонність праці, емоційні перевантаження). 6 Соціальні НШВФ – це неякісна організація роботи, понаднормова робота, змушеність праці в колективі з поганими відносинами між його членами,

соціальна ізольованість з відривом від сім_ї, зміна біоритмів, незадоволеність роботою, фізична та/або словесна образа та її ризик, насильство та його ризик. Один і той же НШВФ за природою своєї дії може належати водночас до різних груп.

4.2 Вплив вібрації на людину

Вібрація - це механічні коливання пружних тіл або коливальні рухи механічних систем. Для людини вібрація є видом механічного впливу, який має негативні наслідки для організму [34].

Причиною появи вібрації є неврівноважені сили та ударні процеси в діючих механізмах. Створення високопродуктивних потужних машин і швидкісних транспортних засобів при одночасному зниженні їх матеріалоємності неминуче призводить до збільшення інтенсивності і розширення спектру вібраційних та віброакустичних полів. Цьому сприяє також широке використання в промисловості і будівництві високоефективних механізмів вібраційної та віброударної дії.

Дія вібрації може приводити до трансформування внутрішньої структури і поверхневих шарів матеріалів, зміни умов тертя і зносу на контактних поверхнях деталей машин, нагрівання конструкцій. Через вібрацію збільшуються динамічні навантаження в елементах конструкцій, стиках і сполученнях, знижується несуча здатність деталей, ініціюються тріщини, виникає руйнування обладнання. Усе це приводить до зниження строку служби устаткування, зростання імовірності аварійних ситуацій і зростання економічних витрат. Вважають, що 80% аварій в машинах і механізмах здійснюється внаслідок вібрації. Крім того, коливання конструкцій часто є джерелом небажаного шуму. Захист від вібрації є складною і багатоплановою в науково-технічному та важливою у соціальноекономічному відношеннях проблемою нашого суспільства [34].

Вплив вібрації на людину залежить від її спектрального складу, напрямку дії, прикладення, тривалості впливу, а також від індивідуальних особливостей людини. При оцінці вібраційного впливу потрібно враховувати, що коливальні

процеси притаманні живому організму. В основі серцевої діяльності і кровообігу та біострумів мозку лежать ритмічні коливання. Внутрішні органи людини можна розглядати як коливальні системи з пружними зв'язками. Частоти їх власних коливань лежать у діапазоні 3..6 Гц. Частоти власних коливань плечового пояса, стегон і голови щодо опорної поверхні (положення стоячи) складають 4...6 Гц, голови щодо плечей (положення сидячи) 25...30 Гц.

При впливі на людину зовнішніх коливань (хитавиці, струсів, вібрації) відбувається їхня взаємодія з внутрішніми хвильовими процесами, виникнення резонансних явищ. Так, зовнішні коливання частотою менш 0,7 Гц утворюють хитавицю і порушують у людини нормальну діяльність вестибулярного апарата. Інфразвукові коливання (менш 16 Гц), впливаючи на людину, пригнічують центральну нервову систему, викликаючи почуття тривоги, страху. При певній інтенсивності на частоті 6..7 Гц інфразвукові коливання, втягуючи у резонанс внутрішні органи і систему кровообігу, здатні викликати травми, розриви артерій, тощо [34].

Вібрація, що діє на людину, має широкий діапазон – від десятих часток одного до декількох тисяч Гц. Характерними ознаками шкідливого впливу вібрації на людину є можливі зміни у функціональному стані: підвищена втома, збільшення часу моторної реакції, порушення вестибулярної реакції. Медичними дослідженнями встановлено, що вібрація є подразником периферичних нервових закінчень, розташованих на ділянках тіла людини, що сприймають зовнішні коливання. Адекватним фізичним критерієм оцінки її впливу на організм людини є коливальна енергія, що виникає на поверхні контакту, а також енергія, поглинена тканинами і передана опорно-руховому апарату та іншим органам. У результаті впливу вібрації виникають нервовосудинні розлади, ураження кістково-суглобної та інших систем організму. Відзначаються, наприклад, зміни функції щитовидної залози, сечостатевої системи, шлунково-кишкового тракту. Так, медичні дослідження показали , що у працюючих в умовах вібрації відбуваються значні зміни кістковосуглобної системи, які виражаються у функціональній перебудові кісткової тканини, регіональному остеопорозі, кістовидних утвореннях у кістках, асептичному некрозі кісток, хронічних

переломах. Відзначається, що терміни виникнення змін у кістках у працівників вібраційних професій коливається в межах від 6-8 місяців до 2-5 років.

Шкідливість вібрації збільшується при одночасному впливі на людину таких факторів, як знижена температура, підвищений шум, запиленість повітря, тривала статична напруга тощо. Сучасна медицина розглядає виробничу вібрацію як могутній стрес-фактор, що має негативний вплив на психомоторну працездатність, емоційну сферу і розумову діяльність, підвищує ймовірність виникнення різних захворювань і нещасних випадків. Особливо небезпечний тривалий вплив вібрації для жіночого організму. Широкий комплекс патологічних відхилень, викликаний впливом вібрації на організм людини, кваліфікується як віброзахворювання [34].

Вібрація як фізичний чинник виробничого середовища спостерігається в металообробній, гірничодобувній, металобудівній, машинобудівній, авіаційній та інших галузях народного господарства. Джерелом вібрації можуть бути різні механізми, вібраційне устаткування, віброінструменти, акустичні системи, транспортні та сільськогосподарські машини.

Загальна вібрація поділяється на транспортну вібрацію, яка діє на людину на робочих місцях в транспортних засобах (трактори сільськогосподарські та промислові, самохідні сільськогосподарські машини (комбайни), тягачі, грейдери ті інші); транспортно-технологічну вібрацію, яка діє на людину на робочих місцях машин з обмеженою рухливістю (екскаватори, крани промислові та будівельні, гірничі комбайни, транспорт виробничих приміщень та інші) та технологічну вібрацію, яка діє на людину на робочих місцях стаціонарних машин чи передається на робочі, де немає джерел вібрації (верстати та метало-деревообробне, пресувально-ковальське обладнання, ливарні машини, електричні машини, насосні агрегати та вентилятори, обладнання для буріння свердловин, бурові верстати, машини для тваринництва, очищення та сортування зерна (у тому числі сушарні), обладнання промисловості будматеріалів (крім бетоноукладачів), установки хімічної та нафтохімічної промисловості та інші.

Оператори машин, які зазнають у процесі трудової діяльності впливу вібрації, підлягають попереднім та періодичним медичним оглядам відповідно

до Порядку проведення медичних оглядів працівників певних категорій, затвердженого Наказом МОЗ України від 21.05.2007 р. №246. Обов'язкові попередні (під час прийняття на роботу) та періодичні (протягом трудової діяльності) медичні огляди дозволять визначити стан здоров'я працівника та можливість виконання без погіршення стану здоров'я професійних обов'язків, своєчасно виявити ранні ознаки хронічного професійного захворювання, забезпечує динамічне спостереження за станом здоров`я в умовах дії шкідливих та небезпечних факторів і трудового процесу, вирішує питання щодо можливості продовжувати роботу в умовах дії шкідливих та небезпечних факторів і трудового процесу [34].

За результатами періодичних медичних оглядів роботодавець забезпечує проведення відповідних оздоровчих заходів Заключного акта у повному обсязі та усуває причини, що призводять до професійних захворювань. Організовує проведення лабораторних досліджень умов праці на робочих місцях та вживає заходів до усунення небезпечних і шкідливих для здоров'я виробничих факторів.

До роботи операторами машин допускаються особи не молодші 18 років, які пройшли попередній медичний огляд, мають відповідну кваліфікацію та ознайомлені з характером впливу вібрації на організм.

ВИСНОВКИ

У роботі запропоновано модель відновлення зашумленого сигналу на основі BLSTM з IBM маскою на виході.

Мережа навчалася та тестувалася на наборі даних TIMIT та показала недостатню ефективність. Далі модель була модифікована додаванням додаткового етапу кластеризації даних. Було розглянуто три методи кластеризації: K-Means, Mean-Shift, EM. Для кожної моделі було обчислено метрики ERLE, PESQ, STOI, що характеризують її якість. Використання алгоритмів кластеризації EM Mean-Shift виявилося неефективним порівняно з методом BLSTM при співвідношенні сигнал/ехо 10 дБ. При співвідношенні сигнал/ехо 6 дБ BLSTM+Mean-Shift призвело до незначного поліпшення метрики PESQ порівняно з методом BLSTM.

Використання методу K-Means призвело до суттєвого покращення всіх показників, на відміну від методів Mean-Shift, EM. У сценаріях з подвійною розмовою, при співвідношенні сигнал/ехо 10 дБ метрика STOI, що характеризує розбірливість мови, покращилася на 7%, а метрика PESQ, що характеризує якість відновлення мови, на 18.8%.

Надалі запропонована модель BLSTM+k-Means буде використана для завдання придушення акустичного еха за наявності шуму та нелінійних спотворень.

ПЕРЕЛІК ДЖЕРЕЛ

1. Benesty J., Jensen J., Christensen M., Chen J. Speech Enhancement: A Signal Subspace Perspective. Elsevier Academic Press, 2014. 129 p.

2. Lee C.M., Shin J.W., Kim N.S. DNN-based residual echo suppression // Interspeech 2015, Dresden, Germany, September6–10, 2015. ISCA, 2015. P. 1775 – 1779.

3. Zhang H., Wang D. Deep learning for acoustic echo cancellation in noisy and double-talk scenarios // Interspeech 2018, Hyderabad, India, September 2-6, 2018. ISCA, 2018. P. 3239-3243.

4. Zhang H., Tan K., Wang D. Deep learning for joint accoustic echo and noise cancellation with nonlinear distortions // Interspeech 2019, Graz, Austria, September 15-19, 2019. ISCA, 2019. P. 4255-425.

5. Wang D. On Ideal Binary Mask є Computational Goal of Auditory Scene Analysis // Speech Separation by Humans and Machines / ed. by P. Divenyi. Springer, Boston, MA, 2005. P. 181-197.

6. Li N., Loizou PC Factors influencing intelligibility of ideal binary- masked speech: Implications for noise reduction // J. Acoust. Soc. Am. 2008. Vol. 123, no. 3. P. 1673-1682.

7. Brungart D.S., Chang P.S., Simpson B.D., Wang D. Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation // J. Acoust. Soc. Am. 2006. Vol. 120, no. 6. P. 4007–4018.

8. Benesty J., G'ansler T., Morgan DR, et al. Advances in network and acoustic echo cancellation.

9. Enzner G., Buchner H., Favrot A., Kuech F. Chapter 30 - Acoustic Echo Control // Academic Press Library in Signal Processing: Volume 4 / ed. by J. Trussell, A. Srivastava, AK Roy-Chowdhury, et al. ELSEVIER, 2014. P. 807-877.

10. Hamidia M., Amrouche A. A new robust double-talk detector based on the Stockwell transform for acoustic echo cancellation // Digital Signal Processing. 2017. Vol. 60. P. 99–112,

11. Ykhlef F., Ykhlef H. Post-filter for acoustic echo cancellation in frequency domain // 2014 Second World Conference on Complex Systems (WCCS), Agadir, Maroko, Nov 10-12, 2014. IEEE, 2014. P. 446-4

12. Kuech F., Kellermann W. Nonlinear residual echo suppression using a power filter model of acoustic echo path //2007 International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, Honolulu, HI, USA, i April 15-20, 70.202.

13. Malek J., Koldovsk'y Z. Hammerstein model-based nonlinear echo cancelation using a cascade of neural network and adaptive linear filter // 2016 IEEE International Workshop on Acoustic Signal Enhancement (IWAENC), Xi'an, China, Sept 13–16, 2016. IEEE, 2016. P. 1–5.

14. Yang F., Wu M., Yang J. Stereophonic acoustic echo suppression based on wiener filter in the short-time fourier transform domain // IEEE Signal Processing Letters. 2012. Vol. 19, no. 4. P. 227–230.

15. Wang D., Chen J. Supervised speech separation based on deep learning: overview // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2018. Vol. 26, no. 10. P. 1702-1726.

16. Wang Y., Narayanan A., Wang D. On training targets for supervised speech separation // IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2014. Vol. 22, no. 12. P. 1849–1858.

17. Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997. Vol. 9, no. 8. P. 1735-1780.

18. Erdogan H., Hershey JR, Watanabe S., Roux JL Phase-sensitive and recognition-boosted speech separation using deep recurrent neural networks //2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane 2 2015. P. 708-712.

19. Weninger F., Erdogan H., Watanabe S., et al. Speech Enhancement with LSTM Recurrent Neural Networks and its Application to Noise-Robust ASR // Latent Variable Analysis and Signal Separation. Vol. 9237 / ed. by E. Vincent, A. Yeredor, Z. Koldovsk'y, P. Tichavsk'y. Cham: Springer International Publishing, 2015. P. 91–99.

Lecture Notes in Computer Science.

20. Chen J., Wang D. Long short-term memory for speaker generalization in supervised speech separation // The Journal of Acoustical Society of America. 2017. Vol. 141, no. 6. P. 4705-4714.

21. Жуковський, В. В., Шатний, С. В., & Жуковська, Н. А. (2020). Нейронна мережа для розпізнавання та класифікації картографічних зображень ґрунтових масивів. Scientific Bulletin of UNFU, 30(5), 100-104. https://doi.org/10.36930/40300517.

22. Shumaila MN A Comparison of K-Means and Mean Shift Algorithms // International Journal of Theoretical and Applied Mathematics. 2021. Vol. 7, no. 5. P. 76-84.

23. Palmqvist M. Methods and algorithms for quality and performance evaluation of audio conferencing systems: PhD thesis / Palmqvist Maria. Umer a University, Faculty of Science, Technology, Department of Physics, Sweden, 2013..

24. ITU-T. Recommendation P.862, Perceptual Evaluation of Speech Quality (PESQ). 2001.

25. Fu S.-W., Liao C.-F., Tsao Y. Learning with Learned Loss Function: Speech Enhancement with Quality-Net for Improve Perceptual Evaluation of Speech Quality // EEE Signal Processing Letters. 2020. Vol. 27. P. 26-30.

26. Zermini A. Deep Learning for Speech Separation: PhD thesis / Zermini Alfredo. University of Surrey, Faculty of Engineering, Physical Sciences, Centre for Vision, Speech, Signal Processing (CVSSP), South East of England, UK, 2020.

27. Xia S., Li H., Zhang X. Using Optimal Ratio Mask as Training Target for Supervised Speech Separation // 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), Kuala Lumpur, Malaysia, Dec 12–15, 2017. IEEE, 2017. P. 163–166.

28. Allen J.B., Berkley D.A. Image method for efficiently simulating small-room acoustics // The Journal of the Acoustical Society of America. 1998. Vol. 65, no. 4. P. 943–950.

29. Vorlaender M. Auralization: fundamentals acoustics, modelling, simulation, algorithms and acoustic virtual reality. Berlin: Springer-Verlag,2008.340p.

30. Schroeder D. Physically based real-time auralization of interactive virtual environments: PhD thesis / Schroeder Dirk. RWTH Aachen University, 2011.

31. Reverberation. [Електронний ресурс] – Режим доступа: https://www.sciencedirect.com/topics/physics-and-astronomy/reverberation (дата звернення: 05.05.2025).

32. Tensorflow. [Електронний ресурс] – Режим доступа: http://www.tensorflow.org/. (дата звернення: 05.05.2025).

33. Заікіна Д., Глива В. Основи охорони праці та безпека життєдіяльності. 2019. URL: https://doi.org/10.31435/rsglobal/001 (дата звернення: 14.04.2025).

34. Безпека в надзвичайних ситуаціях. Методичний посібник для здобувачів освітнього ступеня «магістр» всіх спеціальностей денної та заочної (дистанційної) форм навчання / укл.: Стручок В. С. Тернопіль: ФОП Паляниця В. А., 2022. 156 с.

ДОДАТКИ

Додаток А

Функції для отримання вхідних та вихідних даних моделі BLSTM (фрагмент)

```
import numpy as np
import librosa
import pickle
import os
import matplotlib.pyplot as plt
import time
import librosa.display
from scipy.io import wavfile

class DataGenerator(object):

    def __init__(self):

        self.mean = -1.681676
        self.std = 0.768120

    def stft(self, wavfile, frame_size):
        stft = librosa.stft(y=wavfile, n_fft=frame_size, window='hann')
        mag, phase = librosa.magphase(stft)
        stft = np.transpose(stft)
        mag = np.transpose(mag)
        phase = np.transpose(np.angle(phase))
        return stft, mag, phase

    def istft(self, stft_matrix):
        stft_matrix = np.transpose(stft_matrix)
        istft = librosa.istft(stft_matrix=stft_matrix, window='hann')
        return istft

    def GetListOfFiles(self, data_dir):

        speakers_dir = [os.path.join(data_dir, i) for i in os.listdir(dat
a_dir)]
        n_speaker = len(speakers_dir)
        speaker_file = {}
        for i in range(n_speaker):

            wav_dir_i = [os.path.join(speakers_dir[i], file) for file in
os.listdir(speakers_dir[i])]
            if i not in speaker_file:
                speaker_file[i] = []
            for j in wav_dir_i:
                if j.endswith(".WAV"):
                    speaker_file[i].append(j)

        return n_speaker, speaker_file

    def GetSeperateSignals(self, wavfile1, wavfile2, sampling_rate):
        speech_1, _ = librosa.core.load(wavfile1, sr=sampling_rate)
        audio, _ = librosa.core.load(wavfile2, sr=sampling_rate)

        rt60 = 0.5
```

Додаток Б

Вихідний код для побудови та навчання BLSTM (фрагмент)

```
import time
batch_size = 64;
tr_features = np.concatenate([item['SampleStd'] for item in data], axis=0
)
Y_labels = (np.concatenate([np.asarray(item['IBM'])[:,:,0] for item in da
ta], axis=0)).astype(int)
n_features = np.shape(tr_features)[1]
n_classes = np.shape(Y_labels)[1]
validation_list = ['/gdrive/My Drive/Data1_1/test' + str(i) + '.pkl' for
i in range(1, 2)]
validation_data_generator = DataGenerator2(validation_list)
validation_batches = int(validation_data_generator.total_batches(batch_si
ze))
print("Training set: Data samples: %d, Total number of points: %d"%(valid
ation_data_generator.total_samples, validation_data_generator.total_numbe
r_of_datapoints()))
tf.reset_default_graph()
training_epochs = 100
n_neurons_in_h1 = 300
n_neurons_in_h2 = 300
learning_rate = 0.01
frames_per_sample = 100
embedding_dimension = 3

X = tf.placeholder(tf.float32, shape=[None, frames_per_sample, n_features
], name="features")
Y = tf.placeholder(tf.float32, shape=[(batch_size * frames_per_sample), n
_classes, 2], name="labels")
with tf.variable_scope('BLSTM1') as scope:
    lstm_fw_cell1 = tf.contrib.rnn.BasicLSTMCell(n_neurons_in_h1)

    lstm_bw_cell1 = tf.contrib.rnn.BasicLSTMCell(n_neurons_in_h1)
    outputs1, states1 = tf.nn.bidirectional_dynamic_rnn(cell_fw=lstm_fw_c
ell1, cell_bw=lstm_bw_cell1, inputs=X, dtype=tf.float32)
    state_concate1 = tf.concat(outputs1, 2)

with tf.variable_scope('BLSTM2') as scope:
    lstm_fw_cell2 = tf.contrib.rnn.BasicLSTMCell(n_neurons_in_h2)
    lstm_bw_cell2 = tf.contrib.rnn.BasicLSTMCell(n_neurons_in_h2)

    outputs2, states2 = tf.nn.bidirectional_dynamic_rnn(cell_fw=lstm_fw_c
ell2, cell_bw=lstm_bw_cell2, inputs=state_concate1, dtype=tf.float32)
    state_concate2 = tf.concat(outputs2, 2)
    out_concate = tf.reshape(state_concate2, [-1, n_neurons_in_h2 * 2])

W0 = tf.Variable(tf.random_normal([2 * n_neurons_in_h2, n_classes * embed
ding_dimension], mean=0, stddev=1/np.sqrt(n_features)), name="weightsOut"
)
b0 = tf.Variable(tf.random_normal([n_classes * embedding_dimension], mean
=0, stddev=1/np.sqrt(n_features)), name="biasesOut")
a = tf.nn.sigmoid((tf.matmul(out_concate, W0)+b0), name="activationOutput
Layer")
reshaped_emb = tf.reshape(a, [-1, n_classes, embedding_dimension])
normalized_emb = tf.nn.l2_normalize(reshaped_emb, 2)
```