

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет інформаційних радіотехнологій та технічного захисту інформації
(повна назва)

Кафедра медіаінженерії та інформаційних радіоелектронних систем
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження методів режекції голосу при дубляжі фільмів.
(тема)

Виконав:
студент 2 курсу, групи МІМ-22-1
Чернов М.М.
(прізвище, ініціали)

Спеціальність 172 Телекомунікації та
радіотехніка
(код і повна назва спеціальності)

Тип програми освітньо-професійна

Освітня програма Медіаінженерія
(повна назва освітньої програми)

Керівник доц. Посошенко В.О.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри Володимир КАРТАШОВ
(підпис)

2023 р.

Харківський національний університет радіоелектроніки

Факультет інформаційних радіотехнологій та технічного захисту інформації

Кафедра медіаінженерії та інформаційних радіоелектронних систем

Рівень вищої освіти другий (магістерський)

Спеціальність 172 Телекомунікації та радіотехніка
(код і повна назва)

Тип програми освітньо-професійна

Освітня програма Медіаінженерія
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

«_____» _____ 20__ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту Чернову Микиті Максимовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження методів режекції голосу при дубляжі фільмів.

затверджена наказом по університету від " 20 " 10 2023 р. № 1224 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 10.01.2024 р.

3. Вихідні дані до роботи Виконати теоретичну оцінку факторів, що впливають на якість режекції голосу. Визначити головні фактори, що впливають на якісну ізоляцію голосу від інструментальної доріжки. Провести дослідження багатьох методів режекції, як сучасних, так і класичних. Отримати теоретичні оцінки різних методів та їх ефективності. Провести уточнення теоретичних результатів шляхом експериментального дослідження. Провести експериментальне дослідження режекції голосу і створення інструментальної доріжки з мінімальною кількістю спотворень у спектрі.

4. Перелік питань, що потрібно опрацювати в роботі _____

Вступ

1 Аналітичний огляд методів розділення джерел звуку в міксі.

2 Теоретичний опис ефективних методів режекції голосу та основні класифікації.

3 Експериментальне дослідження різних методів режекції голосу для створення інструменталу.

Висновки

Перелік посилань

Додатки

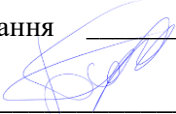
5. Перелік графічного матеріалу із зазначенням обов'язкових креслеників, схем, плакатів, комп'ютерних ілюстрацій

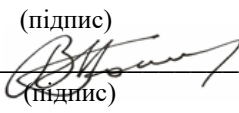
1. Режекція голосу з міксу (1 аркуш А4).
2. Постановка задачі (1 аркуш А4).
3. Шумозаглушення (1 аркуш А4).
4. Активне шумозаглушення (1 аркуш А4).
5. Еквалізація (1 аркуш А4).
6. Частотні діапазони звукових сигналів (1 аркуш А4).
7. Саунд дизайн (1 аркуш А4).
8. Stereo Imager (1 аркуш А4).
9. Нейронні мережі (1 аркуш А4).
10. Порівняння методів (1 аркуш А4).
11. Відео демонстрація (1 аркуш А4).

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналітичний огляд літератури	01.09.23–27.09.23	
2	Теоретичний аналіз методів режекції	28.09.23–11.10.23	
3	Підготовка аудіофрагментів	12.10.23–10.11.23	
4	Експериментальна частина	11.11.23–03.12.23	
5	Обробка результатів	04.12.23–17.12.23	
6	Графічна частина роботи	18.12.23–17.12.23	
7	Перевірка керівником	18.12.23–30.12.23	
8	Перевірка на академічний плагіат	02.01.24–05.01.24	
9	Перевірка завідувачем кафедри, рецензування	06.01.24–09.01.24	

Дата видачі завдання  20.10.2023 р.

Студент  Микита ЧЕРНОВ

Керівник роботи  Віталій ПОСОШЕНКО

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 49 сторінок, 20 рисунків, 1 таблиця, 21 джерело.

ЕКВАЛІЗАЦІЯ, АЧХ, СПЕКТРОГРАМА, ТЕМБР, ЧАСТОТА, СИГНАЛ

Об'єкт дослідження – процес режекції голосу з аудіоміксу для створення інструментальної доріжки.

Мета роботи – розробити найефективнішу методику подавлення голосу у звукозаписі з мінімальною кількістю спотворень у сигналі; визначення можливостей і обмежень автоматизованої системи корекції.

В роботі виконано теоретичну оцінку факторів, що впливають на якість режекції голосу. Визначено головні фактори, що впливають на якісну ізоляцію голосу від інструментальної доріжки. Проведено дослідження багатьох методів режекції, як сучасних, так і класичних. Отримано теоретичні оцінки різних методів та їх ефективності. Проведено уточнення теоретичних результатів шляхом експериментального дослідження. Проведено експериментальне дослідження режекції голосу і створення інструментальної доріжки з мінімальною кількістю спотворень у спектрі.

ABSTRACT

Explanatory note to the qualification work: 49 pages, 20 figures, 1 table, 21 sources.

EQUALIZATION, FREQUENCY RESPONSE, SPECTROGRAM, TIMBRE, FREQUENCY, SIGNAL

The object of study is the process of voice rejection from an audio mix to create an instrumental track.

The purpose of the study is to develop the most effective method of voice suppression in audio recordings with a minimum amount of distortion in the signal; to determine the capabilities and limitations of an automated correction system.

The theoretical evaluation of factors affecting the quality of voice rejection is performed. The main factors affecting the quality of voice isolation from an instrumental track are identified. A study of many rejection methods, both modern and classical, was conducted. Theoretical estimates of various methods and their effectiveness are obtained. The theoretical results were refined by means of an experimental study. An experimental study of voice rejection and the creation of an instrumental track with a minimum amount of distortion in the spectrum was conducted.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів.....	8
Вступ.....	9
1 АНАЛІТИЧНИЙ ОГЛЯД МЕТОДІВ РОЗДІЛЕННЯ ДЖЕРЕЛ ЗВУКУ В МІКСІ.....	11
1.1 Обробка з частотною характеристикою яка змінюється у часі	11
1.2 Шумозаглушення.....	12
1.2.1 Активне шумозаглушення.....	14
1.3 Ізоляція голосу.....	15
1.4 Звуковий тракт і його параметри.....	16
1.5 Частотний аналіз звуку.....	18
1.5.1 Еквалайзери.....	18
1.5.2 Частотний діапазон звукових сигналів.....	21
1.5.3 Амплітудно-частотні характеристики фільтрів.....	26
1.5.4 Компресія.....	27
1.6 Висновки по розділу 1.....	28
2 ТЕОРЕТИЧНИЙ ОПИС ЕФЕКТИВНИХ МЕТОДІВ РЕЖЕКЦІЇ ГОЛОСУ ТА ОСНОВНІ КЛАСИФІКАЦІЇ	29
2.1 Саунд дизайн	29
2.1.1 Історія саунд дизайну.....	29
2.1.2 Запис звуку	31
2.1.3 Цифрові технології	33
2.1.4 Кіновиробництво... ..	34
2.2 Стереоімаджер.....	35
2.3 Нейронні мережі.....	36
2.3.1 Штучні нейронні мережі.....	36
2.3.2 Тренування нейронних мереж.....	36
2.3.5 Кероване навчання.....	37

2.4 Висновки по розділу 2.....	38
3 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ РІЗНИХ МЕТОДІВ РЕЖЕКЦІЇ ГОЛОСУ ДЛЯ СТВОРЕННЯ ІНСТРУМЕНТАЛУ.....	39
3.1 Таблиця дослідження та порівняння.....	39
3.2 Висновки по розділу 3.....	43
Висновки.....	44
Перелік джерел посилань.....	47
ДОДАТКИ.....	50
Додаток А. Графічний матеріал.....	51
Додаток Б. Відомість кваліфікаційної роботи.....	62

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,
СКОРОЧЕНЬ І ТЕРМІНІВ

- АЧХ – амплітудно-частотна характеристика;
ВЧ – верхні частоти;
ІНСТ – інструментальна доріжка;
UVR – ultimate vocal remover;
дБ – децибели;
ЧД – частотний діапазон;
НЧ – нижні частоти;
КОМП – компресор;
М – мікрофон;
МС – мовний сигнал;
ЕК – еквайзер;
ОТ – основний тон;
Пр – приміщення студії;
СГ – спектрограма;
ШІ – Штучний Інтелект
Frequency – центральна (робоча) частота;
Gain – підсилення чи ослаблення обраної смуги.

ВСТУП

Кіно – це аудіовізуальне мистецтво. У цьому значенні 50% терміну займає слово аудіо. Звук із головою занурює глядача у те, що відбувається на екрані. Поява звуку в кіно радикально змінила трактування екранного часу та простору. Вже у перших звукових фільмах можна знайти як приклади створення єдиного просторово-часового континууму з допомогою звуку, а й "розширення" екранного простору рахунок звукового тла при мінімальному використанні образотворчих засобів. Ймовірно, цей прийом швидко прижився в кіномистецтві з тієї причини, що принципи створення звукового простору на той час були успішно освоєні в радіопостановках (зміна крупності звукового плану; створення ефекту реверберації; позначення середовища, де відбувається дія, за допомогою шумів).

У сучасному світі, де кінематографічне мистецтво та технологічні досягнення взаємодіють, проблема видалення голосу зі звукової доріжки фільмів набуває особливої актуальності. Здатність розділити аудіо складові фільму є важливим кроком у розвитку кінематографу, дозволяючи створювати альтернативні аудіо-версії, використовувати озвучення різних мов, або виправляти акустичні дефекти під час постпродакшну.

В результаті цього дослідження очікується, що нові методи режекції голосу зі звукової доріжки фільму, серіалу чи аніме зможуть відкрити більше простору для розвитку аудіовізуальної індустрії. Це дозволить використовувати альтернативні мови озвучення витворів кіномистецтва та залучити більш широку аудиторію. Крім того, відкриються нові можливості виправлення спотворень в обробленому сигналі.

Крім того дослідження покаже перспективи обробки сигналу та використання нейронних мереж для видалення голосу. Ці технології можуть бути використані не тільки у кінематографі, а також у музиці, телевізійній та рекламних сферах.

В цілому дослідження методів режекції голосу є важливим кроком у розвитку кінематографу та постпродакшену. Використання нових технологій у вже сталих процесах допоможе задовольнити потреби більшого пласту аудиторії та охопити українському дубляжу набагато більше цікавих проектів.

Дослідження в даній роботі відповідають традиційному напрямку наукових робіт колективу кафедри МІРЕС.

1 АНАЛІТИЧНИЙ ОГЛЯД МЕТОДІВ РОЗДІЛЕННЯ ДЖЕРЕЛ ЗВУКУ В МІКСІ

1.1 Обробка з частотною характеристикою, яка змінюється у часі

Обробка з частотною характеристикою яка змінюється у часі – метод обробки аудіосигналу який використовує часово частотний аналіз для зміни частотної характеристики сигналу. Цей підхід дає можливість адаптувати обробку сигналу в залежності від його динаміки та характеристики в певний момент часу.

Основні методи включають наступні перетворення.

Швидке перетворення Фур'є (ШПФ) - це перетворення, яке використовується для визначення синусоїдальної частоти і фази локальних ділянок сигналу при його зміні в часі. На практиці процедура обчислення ШПФ полягає в тому, щоб розділити довший за часом сигнал на коротші сегменти однакової довжини, а потім обчислити перетворення Фур'є окремо для кожного коротшого сегмента. Це показує спектр Фур'є на кожному коротшому відрізьку. Потім зазвичай будують графік зміни спектру як функції часу, відомий як спектрограма або графік водоспаду, який зазвичай використовується в дисплеях спектру на основі програмно визначеного радіо (SDR). Повно діапазонні дисплеї, що охоплюють весь діапазон SDR, зазвичай використовують швидке перетворення Фур'є (ШПФ) з 2^{24} точками на настільних комп'ютерах.

Вейвлет перетворення.

У математиці вейвлет-ряд - це представлення квадратично інтегрованої (дійсної або комплексної) функції певним ортонормованим рядом, породженим вейвлетом. Це допомагає аналізувати сигнал на різних масштабах та частотах. Також таке перетворення допомагає виявити частотні та часові особливості сигналу.

Кепстральний аналіз.

Це інструмент для виявлення періодичності в частотному спектрі, який в основному використовується для визначення висоти польоту, в радіолокаційних і гідролокаційних системах, аналізі мови і діагностиці. Цей метод дозволяє представити сигнал у двовимірному просторі, де одна вісь відповідає за час, а інша за частоту, що допомагає визначити спектральні зміни вздовж усієї частотної вісі

1.2 Шумозаглушення

Шумозаглушення – процес зменшення рівня стороннього шуму в сигналі задля покращення якості звуку. Цей процес може бути використаний в аудіо та відео обробці. Також його можна зустріти у різних аудіо пристроях, таких як навушники, мікрофони, рекордери.

Основні методи шумозаглушення включають у себе такі методи.

Фільтрація за часовими характеристиками. Використання фільтрів які виявляють і фільтрують компоненти шуму базуючись на часових характеристиках. Цей метод добре працює для постійного, або повільно змінюючогося шуму.

Адаптивна фільтрація. Адаптивна фільтрація – це система з лінійним фільтром, що має передатну функцію, керовану змінними параметрами, і засоби для налаштування цих параметрів відповідно до алгоритму оптимізації. Через складність алгоритмів оптимізації майже всі адаптивні фільтри є цифровими фільтрами. Адаптивні фільтри застосовуються коли якісь параметри бажаної операції обробки (наприклад, розташування відбиваючих поверхонь у ревербераційному просторі) невідомі заздалегідь або змінюються. Замкнутий адаптивний фільтр використовує зворотний зв'язок у вигляді сигналу помилки для уточнення своєї передатної функції.

Замкнутий адаптивний процес передбачає використання функції вартості, яка є критерієм оптимальної роботи фільтра, для подачі алгоритму,

який визначає, як модифікувати передатну функцію фільтра, щоб мінімізувати вартість на наступній ітерації. Найпоширенішою функцією вартості є середній квадрат сигналу помилки.

Зі збільшенням потужності цифрових сигнальних процесорів адаптивні фільтри стали поширенішими і зараз регулярно використовуються в таких пристроях, як мобільні телефони та інші комунікаційні пристрої, відеокамери та цифрові фотоапарати.

Методи машинного навчання. Метод машинного навчання для шумозаглушення стає все популярніше завдяки своїй адаптивності до різних шумових сценаріїв та умов. Може включати такі підходи як:

Глибоке навчання для автоматичного вивчення складних сценаріїв між вхідним сигналом та шумом.

Генеративні моделі з використанням генеративних архітектур які є згенерованими сценаріями шумових шаблонів.

Адаптивні фільтри які навчають адаптуватись до змін умов шумового сигналу в реальному часі.

1.2.1 Активне шумозаглушення

Активне шумозаглушення (АШЗ), також відоме як шумозаглушення (ШЗ) або активне зменшення шуму (АЗШ), - це метод зменшення небажаних звуків шляхом додавання другого звуку, спеціально призначеного для приглушення першого. Вперше концепція була розроблена наприкінці 1930-х років; подальші розробки, що розпочалися в 1950-х роках, врешті-решт призвели до того, що наприкінці 1980-х років ця технологія стала доступною в комерційних авіаційних гарнітурах. Технологія також використовується в дорожніх транспортних засобах, мобільних телефонах і навушниках.

Звук – це хвиля тиску, яка складається з періодів стиснення і розрідження, що чергуються. Динамік з шумозаглушенням випромінює звукову хвилю з тією ж амплітудою, але з інвертованою фазою (також

відомою як протифаза) по відношенню до оригінального звуку. Хвилі об'єднуються, утворюючи нову хвилю, в процесі, який називається інтерференцією, і ефективно компенсують одна одну – ефект, який називається деструктивною інтерференцією (рис. 1.1).

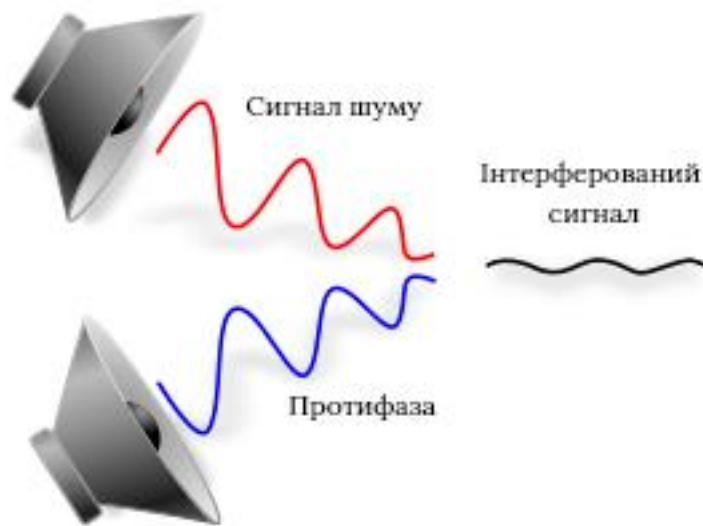


Рисунок 1.1 – Активне шумоподавлення

Сучасний активний контроль шуму, як правило, досягається за допомогою аналогових схем або цифрової обробки сигналів. Адаптивні алгоритми призначені для аналізу форми сигналу фонового звукового або незвукового шуму, а потім на основі певного алгоритму генерують сигнал, який або зсуває фазу, або інвертує полярність вихідного сигналу. Цей інвертований сигнал (в протифазі) потім посилюється і перетворювач створює звукову хвилю, прямо пропорційну амплітуді вихідного сигналу, створюючи деструктивну інтерференцію. Це ефективно зменшує гучність сприйманого шуму.

Динаміки з функцією шумозаглушення можуть розташовуватися поруч із джерелом звуку, який потрібно приглушити. У цьому випадку вони повинні мати той самий рівень звукової потужності, що й джерело небажаного звуку, щоб придушити шум. Альтернативно, датчик, що випромінює сигнал відміни, може бути розташований в місці, де потрібно

послабити звук (наприклад, біля вуха користувача). Це вимагає набагато нижчого рівня потужності для шумозаглушення, але є ефективним лише для одного користувача. Придушення шуму в інших місцях складніше, оскільки тривимірні хвильові фронти небажаного звуку і сигналу придушення можуть збігатися і створювати зони конструктивних і деструктивних перешкод, що чергуються, зменшуючи шум в одних місцях і подвоюючи шум в інших. У невеликих закритих приміщеннях (наприклад, у салоні автомобіля) глобального зменшення шуму можна досягти за допомогою декількох динаміків і мікрофонів зі зворотним зв'язком, а також вимірювання модальних реакцій корпусу.

1.3 Ізоляція голосу

Ізоляція голосу – виділення, або видалення голосової частини сигналу. На виході залишається лише голосова інформація та видаляються усі небажані звуки, або шуми. Цей процес важливий в різних сферах, таких як аудіопродакшн, редагування фільмів, серіалів, аніме та музичному виробництві.

Основні методи ізоляції голосу:

– спектральний аналіз. Він використовує вищезгадані швидкі перетворення Фур'є для розкладання сигналу на частотні компоненти. Також частоти можуть бути визначені та ізольовані базуючись на унікальних характеристиках у спектрі;

– алгоритми розпізнавання голосу. Використовує алгоритми машинного навчання розпізнавання голосу задля ідентифікації та виділення саме голосової інформації;

– сегментація сигналу. Розділяє аудіозапис на сегменти, або блоки та ідентифікує голосову частину.

Задля кращого ефекту процес найчастіше вимагає комбінування усіх методів.

1.4 Звуковий тракт і його параметри

Звуковим трактом - інший пристрій, що виконує передачу і (або) перебудови звуку. Тракт характеризується наступними параметрами:

- номінальний вхідний і вихідний рівень (Input / Output Level) 34;
- міра сигналу на вході і виході тракту, до якого він зберігає зорієнтовані параметри. Вказується в вольтгах і зазвичай приступає за 0 (дБ);
- великий вхідний і святковий рівень - міра сигналу, в який битий шлях охороняє працездатність. Рівні сигналів від номінального до максимального постійно мають ненульовий ґрунтовний рівень;
- показник примноження - розташування величини святкового сигналу до вхідного. Вказується в разгах, відсотках або децибелах;
- спектр частот (Frequency Response) – частотний інтервал, в якому битий шлях охороняє свої генеральні характеристики. Іноді нуль присутність є, що постійний струм;
- форма амплітудно-частотної характеристики (АЧХ) - діаграм підневільності амплітуди сигналу статист від його частоти близько обов'язкової амплітуди сигналу на вході. Биті шляхи з горизонтальним всередині частотним спектром АЧХ називають частотно-незалежним;
- неритмічність АЧХ - відмінності графіка від заданої форми. Вказується у відсотках або децибелах;
- рівень шуму (Noise Level) - міра гулу порівняно номінального рівня сигналу. Вказується в децибелах і постійно має негативне значення. Інша найменування - відповідність меандр / шум (Signal to Noise Ratio, SNR), яке має в своєму розпорядженні настільки ж ґрунтовне значення. Часом орієнтує рівень шуму, занурений до входу - в припущенні, що весь шум надходить виключно для вхід, сам же битий шлях особистого гулу не має;
- показник гармонік (Total Harmonic Distortion, THD) - міра другорядних правильних складових, записуються лінійністю тракту. Вказується у відсотках через величини сигналу; в ряді випадків вказується

для різних гармонік (на говір найбільші заломлення записують непарні гармошки довжелезних порядків);

- рівень інтермодуляційних спотворень (InterModulation Distortion, IMD) - відповідний рівень чужорідних частотних компонентів, породжені 35 обопільною модуляцією корисно утворюють сигналу. Вказується у відсотках через величини сигналу;

- перехідні припинення (Stereo Crosstalk) – ступінь падіння сигналу близько його проникненні в сусідній стереоканал. Вказується в децибелах;

- динамічний спектр (Dynamic Range) – спектр найбільшого і найменшого рівнів сигналу, всередині яких зберігаються генеральні характеристики тракту. Знизу природно недалеким рівнем шуму, зверхньо - загальним рівнем, тому щохвилини рівновелика відповідності меандр / шум, хоч і не лінійність битого шляху в ряді випадків не дозволяє перенести обсяги в даних областях, а це звужує динамічний діапазон.

1.5 Частотний аналіз звуку

1.5.1 Еквалайзери

Процес коригування або зміни амплітуд сигналів з різними частотами у звукових сигналах називається "еквалізацією" (Equalization). Термін цей вказує на вирівнювання відношень амплітуд сигналів на різних частотах.

Еквалайзер – це набір різних типів фільтрів. Існують два основних типи багатосмугових еквалайзерів: графічні і параметричні.

Графічні еквалайзери мають певну кількість регульованих частотних смуг, кожна з яких має фіксовану робочу частоту та постійну ширину смуги навколо робочої частоти. Діапазон регулювання рівня є однаковим для всіх смуг. Зазвичай, найнижча і найвища смуги є фільтрами типу «полочний», тоді як всі інші мають характеристику типу «дзвін».

Октавність еквайзера обчислюється відносно сусідніх частот регулювання:

$$O = \log_2(f_{n+1} / f_n), \quad (1.1)$$

де n – порядковий номер смуги регулювання еквайзера.

Графічні еквайзери, які використовуються у професійних сферах, зазвичай є 1/2- або 1/3-октавними і мають від 15 до 31 смуг на кожен канал; часто вони комплектуються аналізаторами для зручного налаштування. На рис. 1.2 представлено амплітудно-частотну характеристику (АЧХ) і управління 1/2-октавним еквайзером.

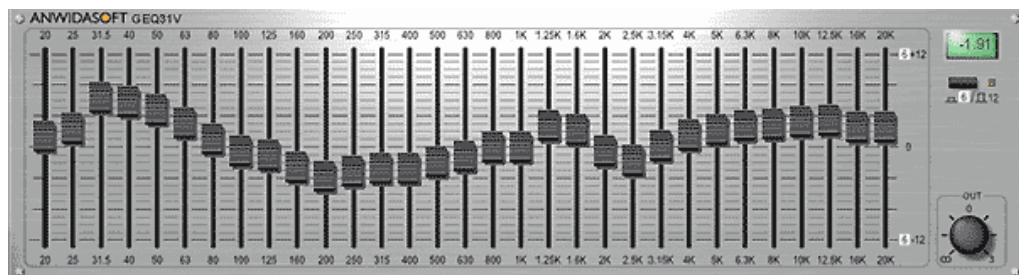


Рисунок 1.2 – Панель управління 1/2-октавного еквайзера

Параметричний еквайзер, завдяки меншій кількості налаштувань, забезпечує значно більші можливості регулювання частотної характеристики сигналу. Кожна смуга має три регульовані параметри:

- Frequency – центральна (робоча) частота (в герцах);
- Q – добротність (налаштовує ширину Δf робочої смуги частот навколо центральної частоти f_0):

$$\Delta f = f_0/Q; \quad (1.2)$$

- Gain – це збільшення чи зменшення амплітуди вибраної смуги (в децибелах).

Таким чином, використання параметричного еквайзера дозволяє більш точно відібрати потрібну частоту і пристосувати смугу та коефіцієнт передачі з більшою точністю.

На рис. 1.2 показано панель керування параметричним еквайзером, а на рис. 1.3 – амплітудно-частотну характеристику при екстремальних положеннях налаштувань Gain, Frequency, Q (добротність).



Рисунок 1.3 – Панель керування параметричного еквайзера

Аналогові параметричні еквайзери застосовуються рідко і характеризуються обмеженою кількістю регульованих смуг частот. Завдяки

досягненням в цифровій обробці звукових сигналів, з'явилися цифрові параметричні еквайзери з практично необмеженим числом регульованих смуг частот. Крім того, цифрові параметричні еквайзери часто мають додаткові параметри, такі як тип фільтру, форма кривої АЧХ і інші. Часто ці еквайзери використовуються як один із блоків обробки в цифрових акустичних процесорах.

Існують еквайзери змішаного типу, які можна знайти в мікшерних консолях. Наприклад, складові НЧ і ВЧ можуть регулюватися аналогічно графічному еквайзеру «поличного» типу, а між ними можуть розташовуватися 2 напівпараметричні смуги (без налаштування добротності).

Також існують гібриди, відомі як параграфічні еквайзери, які представляють собою комбінацію параметричного і графічного еквайзерів. Зазвичай вони дозволяють регулювати величину підсилення (gain) за допомогою повзунків чи на дисплеї, а також мають можливість регулювання добротності та центральної частоти для кожної смуги.

Існують просторові еквайзери, які дозволяють коригувати моно (mid) і стерео (side) складові звукового сигналу окремо. Крім того, існують стереоеквайзери, які надають можливість еквалізувати правий і лівий канали стереосигналів незалежно один від одного.

1.5.2 Частотні діапазони звукових сигналів

Звуковий діапазон можна умовно розділити на вісім чітких частотних смуг, кожна з яких впливає на загальне звучання запису.

Бас, суббас (10 Гц - 80 Гц) – найжнижчі звуки діапазону. Зазвичай вбирають в себе весь бруд голосового запису, гул, шум. Найчастіше просто відрізається у загальному міксі.

Верхній бас (80 Гц - 200 Гц) – це область фундаментальних частот. Несе в собі інформацію про низькі частоти чоловічого голосу, а також іноді є початком частотного діапазону деяких жіночих голосів.



Рисунок 1.4 – Відокремлений бас на параметричному еквайзері



Рисунок 1.5 – Відокремлений верхній бас на параметричному EQ

«Кабінетні частоти» (200 Гц - 500 Гц) – у цьому діапазоні розташовані «бубнящі частоти» жіночого та чоловічого голосів. Також при записі у маленькому приміщенні, реверберація буде проявлятися саме на цьому діапазоні частот.



Рисунок 1.6 – Кабінетні частоти на параметричному EQ

Середні частоти (500 Гц - 1 кГц) – тіло голосу. Саме цей діапазон частот відповідає за і саме вони несуть загальну інформацію про тембр людини.



Рисунок 1.7 – Середні частоти на параметричному EQ

Середні високі (1 кГц - 2 кГц) – «артикуляція». Цей діапазон відповідає за чіткість звучання мови. При підсиленні цього частого діапазону можемо отримати «радіовоїс».



Рисунок 1.8 – Середні високі частоти на параметричному EQ

Презенс (2 кГц - 5 кГц) – високі частоти. Зазвичай несуть в собі інформацію про сибілянти (шиплячі та свистячі звуки). Також дуже часто реверберація великих приміщень припадає саме на цей частотний діапазон.



Рисунок 1.9 – Презенс частоти на параметричному EQ

Високі частоти (5 кГц – 10 кГц) - важливі для створення враження простору у звуці. Вони допомагають розпізнавати різні звукові джерела та їхню просторову взаємодію. Загалом, високі частоти утворюють важливий

елемент аудіопалітри, додаючи вишуканість та виразність до звучання та сприяючи повноті аудіообразу.



Рисунок 1.10 – Високі частоти на параметричному EQ

Аір частоти простору (10 кГц – 20 кГц) – дозволяють чітко виразити гармонічні деталі, такі як металеві відтінки і характеристики атаки і релізу. Цей діапазон важливий для аудіофільського та звукорежисерського сприйняття, а також має вплив на загальне враження від музичного або звукового запису.



Рисунок 1.11 – Високі частоти на параметричному EQ

1.5.3 Амплітудно-частотні характеристики фільтрів

В еквайзерах найчастіше використовуються фільтри з такими характеристиками:

- Low pass – фільтр пропускання низьких частот;
- High pass – фільтр пропускання високих частот;
- Peaking або bell – смуговий фільтр;
- Low shelf – низькочастотний шельф фільтр;
- High shelf – високочастотний шельф фільтр;
- Wahandall Curve – крива нахилу без зміни частот перегину АЧХ.

Можливості цифрових еквайзерів набагато ширші, ніж аналогових, оскільки в них можна моделювати АЧХ за допомогою математичних обчислень, що неможливо зробити в аналоговому вигляді.



Рисунок 1.12 – Low pass та High pass фільтр



Рисунок 1.13 – Peaking фільтр

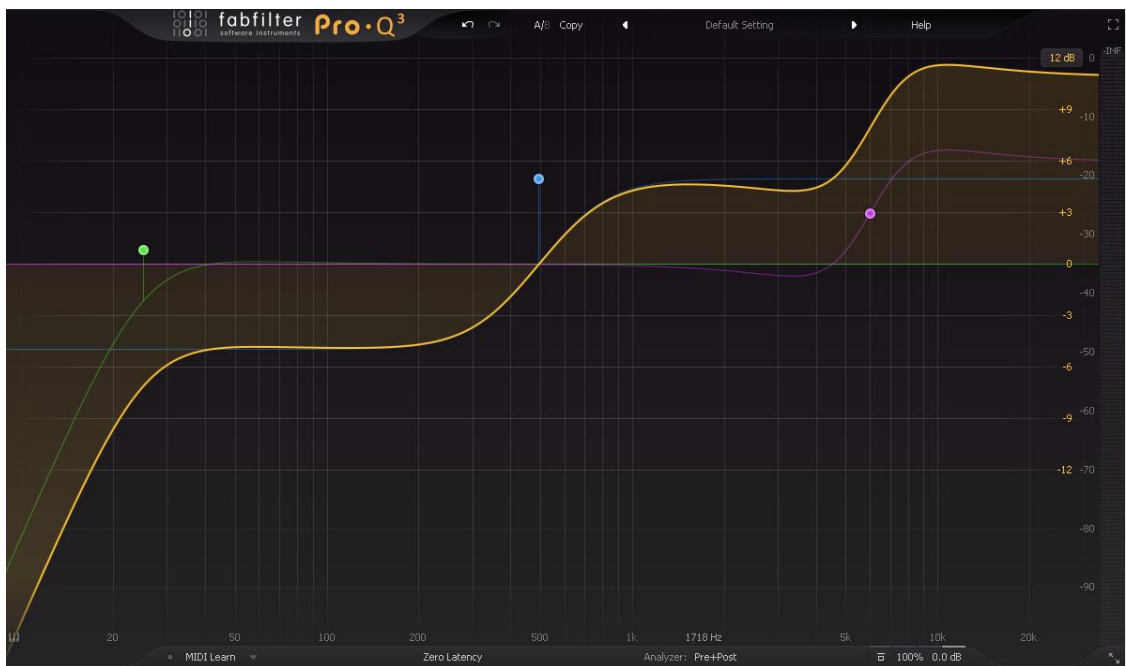


Рисунок 1.14 – Low shelf та High shelf фільтри

1.5.4 Компресія

Компресор – це тип динамічного процесора, який ніби "стягує" динамічний діапазон сигналу і, завдяки цьому, зменшує різницю в рівні гучності між ледве помітними і "піковими" його частинами. Процес

зниження рівня гучності називається ослабленням посилення (gain reduction). Володіючи достатнім досвідом, за допомогою цього процесу можна досягти значно більш щільного звучання.

З цієї причини компресія є найкращим засобом для характеристик, рівень яких змінюється досить широко. Звужуючи динамічний діапазон, компресор підвищує загальний рівень сигналу, не допускаючи спотворень найгучніших його частинах. У той же час, за допомогою компресії можна підтягнути тихі, майже нечутні звуки, такі, як скрип струн і дзвін пружини малого барабану - компресор зробить їх гучнішими, чистішими і набагато помітнішими.

Для налаштування компресора вам зазвичай надаються такі елементи керування:

- Threshold – наскільки гучним повинен бути сигнал до стиснення;
- Ratio – ступінь стиснення. Наприклад, якщо коефіцієнт стиснення встановлений на 6: 1, вхідний сигнал повинен буде перетнути граничне значення на 6 дБ, щоб вихідний рівень збільшився на 1 дБ;
- Attack – як швидко починає працювати компресор;
- Release - через якийсь час після того, як сигнал впаде нижче порогового значення, компресор зупиниться;
- Knee – встановлює, як компресор реагує на сигнали при перевищенні порога. Налаштування Hard Knee означають, що сигнал фіксується відразу, а Soft Knee означає, що компресія включається м'якше, коли сигнал проходить далі граничного значення;
- Make-Up Gain – дозволяє посилити стислий сигнал. оскільки стиснення часто значно послаблює сигнал;
- Outputs – дозволяє підвищити або послабити рівень сигналу компресора.

Завжди є можливість додати компресію в процесі мікшування вже після того, як зроблено запис. Але іноді набагато доречніше використовувати компресор безпосередньо під час запису, і це має свої плюси. По-перше,

сжатый сигнал сразу может продемонстрировать ваши ошибки, что особенно важно при игре на инструменте с широким динамическим диапазоном. По-другому, если мы настраиваем уровень, который «скачет» уже во время записи, это избавит нас от подобных проблем при сведении, сэкономит время.

Количество регулировок компрессора зависит от конструкции, цены и других факторов. Например, приборы на базе VCA имеют по крайней мере пять параметров, которые изменяются: порог срабатывания компрессора, глубина компрессии, время атаки и время восстановления сигнала, а также регулятор выходного уровня. У большинства совершенных моделей встречается вдвое больше параметров, в то время как у большинства компрессоров на оптико-электрических элементах – только две ручки.

1.6 Выводы по разделу 1

Объект исследования – современные методы режески звука для дублирования фильмов.

Цель работы – продемонстрировать перспективы обработки сигнала и использования нейронных сетей для удаления голоса.

Задачи исследования.

1. Провести теоретическую оценку факторов, которые влияют на качество удаления голосового сигнала из микса.
2. Определить основные факторы, которые влияют на создание артефактов во время удаления голосового сигнала.
3. Исследовать современные методы удаления голосового сигнала из микса.
4. Провести экспериментальное исследование по удалению голосового сигнала различными методами.
5. Выделить наиболее эффективный метод удаления голосового сигнала.

2 ТЕОРЕТИЧНЕ ОГЛЯД СУЧАСНИХ МЕТОДІВ РЕЖЕКЦІЇ ЗВУКУ

2.1 Саунд дизайн

Саунд дизайн – один з найефективніших методів видалення голосу з аудіо. Таким чином ми повністю створюємо аудіодоріжку з нуля. Або цей метод можна комбінувати з методами режекції через плагіни, або штучний інтелект. Це мистецтво та практика створення аудіотреків для різноманітних потреб, що включає в себе визначення, отримання або створення звукових елементів за допомогою технік та інструментів аудіопродакшена.

Це використовується в різних галузях, таких як кінематограф, телебачення, розробка відеоігор, театр, звукозапис, відтворення звуку, живе виконання, саунд-арт, пост-продакшн, радіо, нові медіа та розробка музичних інструментів. Зазвичай саунддизайн включає в себе виконання (наприклад, фолія) та редагування раніше створеного або записаного аудіо, такого як звукові ефекти і діалоги для медіа-цілей, але також може передбачати створення звуків з нуля за допомогою синтезаторів. Саунд-дизайнер - це фахівець, який займається цією галуззю.

2.1.1 Історія саунд дизайну

Використання звуку для того, щоб викликати емоції, відобразити настрої і підкреслювати дії у виставах і танцях, почалося ще в доісторичні часи, коли його використовували в релігійних практиках для зцілення або відпочинку. У стародавній Японії в синтоїстських святилищах відбувалися театралізовані дійства під назвою "кагура", які супроводжувалися музикою і танцями.

У середньовіччі п'єси виконувалися у формі театру під назвою комедія дель арте, в якому використовували музику та звукові ефекти для підсилення вистави. Згодом музику і звук почали використовувати в елизаветинському

театрі, де музику і звукові ефекти створювали поза сценою за допомогою таких пристроїв, як дзвони, свистки і роги. У сценарії були написані репліки для музики та звукових ефектів, які повинні були відтворюватися у відповідний час.

У 1913 році італійський композитор Луїджі Русоло розробив механічні апарати для відтворення звуків, які він назвав "інтонарумори". Вони призначалися для використання в футуристичних театральних і музичних виставах. Ці пристрої спроектовані для емуляції різноманітних природних та штучних звуків, таких як шум поїзда або вибух бомби. Трактат Русоло "Мистецтво шумів" став одним із перших письмових документів, що висвітлював використання абстрактних звуків у театрі. Після його смерті його "інтонарумори" використовувалися в більш традиційних театральних постановках для створення реалістичних звукових ефектів.

2.1.2 Запис звуку

Можливо, першим використаним записаним звуком в театрі була фонограф, що відтворювала дитячий плач у лондонському театрі в 1890 р. Шістнадцять років по тому Герберт Бірбом Трі використовував записи у своїй лондонській постановці трагедії Стівена Філліпса "NERO".

Ця подія відзначена в "Театральному журналі" (1906) двома фотографіями: на одній зображений музикант, який сурмить у великий ріг, прикріплений до магнітофона, а на іншій - актор, який записує агонізуючі крики і стогони закатованих мучеників. У статті зазначається: "всі ці звуки реалістично відтворюються грамофоном".

У той час, коли термін "саунд-дизайнер" ще не використовувався, деякі режисери називались "люди ефекти", створюючи і виконуючи звукові ефекти поза сценою, використовуючи поєднання вокальної міміки, механічних та електричних пристроїв і грамофонних платівок. Створенню та виконанню цих ефектів, як натуралістичних, так і абстрактних, приділялося багато уваги.

Протягом двадцятого століття використання записаних звукових ефектів почало витісняти живі звукові ефекти, хоча часто пошук звукових ефектів був обов'язком режисера, а звуковий інженер відтворював записи під час виступів.

2.1.3 Цифрові технології

MIDI та цифрові аудіотехнології сприяли еволюції методів виробництва звуку у 1980-х та 1990-х роках. Цифрові аудіо робочі станції (DAW) та різноманітні алгоритми цифрової обробки сигналів, що застосовуються в них, дозволяють реалізовувати складніші саундтреки з більшою кількістю доріжок та звукових ефектів. Такі функції, як необмежена відміна та редагування на рівні семплів, дозволяють тонко контролювати звукові доріжки.

У сфері театрального звуку функції комп'ютеризованих систем театрального звукового дизайну також були визнані важливими для систем управління живими шоу у Walt Disney World, і, як наслідок, Disney використовував системи такого типу для управління багатьма об'єктами у своєму тематичному парку Disney-MGM Studios, який відкрився у 1989 році. Ці функції були включені в специфікацію MIDI Show Control (MSC), відкритий комунікаційний протокол, який використовується для взаємодії з різними пристроями. Першим шоу, яке повністю використовувало специфікацію MSC, став парад "Чарівне королівство" в "Чарівному королівстві Уолта Діснея" у вересні 1991 року.

Збільшення інтересу до звуку в іграх призвело до розробки більш сучасних інтерактивних аудіоінструментів, які доступні навіть без навичок у комп'ютерному програмуванні. Деякі з таких програм (відомі як "інструменти реалізації" або "аудіо-руші") працюють за аналогією зі звичайними програмами для аудіопродакшену, а також дозволяють звукорежисерам виконувати більш творчі інтерактивні завдання. Ці завдання,

які раніше вимагали втручання програмістів, тепер стали доступнішими. Інтерактивні застосунки також породжують різноманітні методи "динамічного звуку", який взагалі означає звук, що параметрично налаштовується під час виконання програми. Це розширює можливості вираження звуків, наближаючи їх до звуків у фільмах.

Наприклад, звукорежисер може створювати реалістичні та неповторювані звуки кроків, які взаємодіють із візуальним вмістом. Цифрові аудіостанції не можуть безпосередньо "спілкуватися" з ігровими рушіями через непередбачуваний характер подій у грі. У порівнянні з традиційними цифровими аудіостанціями і лінійними медіа (такими, як телебачення чи кіно), де події розгортаються в одному і тому ж порядку, ігри вносять елемент динамічного чи адаптивного зведення.

2.1.4 Кіновиробництво

У кіновиробництві звукорежисер/звукорежисер - це член знімальної групи, відповідальний за весь фільм або окремі його частини. В американській кіноіндустрії посада звукорежисера не контролюється жодною професійною організацією, на відміну від таких посад, як режисер або сценарист.

Терміни "саунд-дизайн" і "саунд-дизайнер" почали використовуватися в кіноіндустрії з 1969 року. Тоді звання звукорежисера було вперше надано Волтеру Мерчу Френсісом Фордом Копполою на знак визнання внеску Мерча у фільм "Люди дощу".

Початкове значення назви "звукорежисер", встановлене Копполою та Мерчем, означало "особа, яка відповідає за всі аспекти звукової доріжки фільму, від запису діалогів та звукових ефектів до перезапису (зведення) фінальної доріжки". Термін "звукорежисер" замінив такі назви, як "звукорежисер" або "мікшер перезапису", для позначення по суті тієї самої посади: головний звукорежисер фінальної звукової доріжки. Такі редактори

та мікшери, як Мюррей Співак ("Кінг-Конг"), Джордж Гровз ("Джазовий співак"), Джеймс Г. Стюарт ("Громадянин Кейн") і Карл Фолкнер ("Подорож до центру Землі") працювали на цій посаді під час студійної ери Голлівуду, і вважались саунд-дизайнерами.

2.2 Стереоімаджер (Stereo Imager)

Цифровий плагін Stereo Imager - це засіб обробки звуку, який використовується для регулювання ширини стереофонічного поля або просторового відображення звуку у міксі (рис. 2.1). Його основна функція полягає в тому, щоб контролювати розташування та роздільність звукових сигналів у просторі між лівим та правим каналами.



Рисунок 2.1 – Плагін Stereo Imager від компанії А.О.М.

Функції які він може виконувати:

- розширення стереобазис: плагін може розширювати стереобазу, тобто відстань між каналами лівого та правого динаміків. Це дозволяє створювати враження простору та ширшого звукового полотна;

- централізація або фокусування: Stereo imager може здійснювати оборот навколо центрального звуку, роблячи його більш чітким та фокусованим. Це особливо корисно для збереження ясності у центральних елементах міксу;

– мід та сайд обробка: плагін може використовувати техніки обробки міді (центральна частина стереозображення) та сайду (бічна частина) для створення більш виразного та збалансованого звучання;

– контроль фази: Stereo imager може впливати на фазові характеристики сигналу для досягнення бажаного ефекту розширення або фокусування.

– візуалізація стереозображення: багато плагінів stereo imager мають графічні інтерфейси, що відображають стерео-просторові характеристики сигналу, щоб користувач міг візуально контролювати зміни.

2.3 Нейронні мережі

2.3.1 Штучні нейронні мережі

Штучні нейронні мережі (ШНМ), часто відомі як нейронні мережі (НМ) або нейромережі, представляють собою обчислювальні системи, які виникли внаслідок натхнення від біологічних нейронних мереж, що утворюють мозок тварин.

Структура ШНМ базується на наборі з'єднаних вузлів, які називаються штучними нейронами. Ці нейрони намагаються апроксимувати функцію біологічних нейронів у мозку. Кожне з'єднання, аналогічно синапсам у біологічному мозку, може передавати сигнали іншим нейронам. Штучний нейрон отримує сигнали, обробляє їх і може передавати сигнали іншим нейронам, з якими він з'єднаний.

Сигнал у з'єднанні представлений дійсним числом, і вихід кожного нейрона обчислюється нелінійною функцією суми його входів. З'єднання називають ребрами, а нейрони та ребра зазвичай мають вагу, яка коригується під час процесу навчання. Вага визначає силу сигналу на з'єднанні, і нейрон може мати поріг, при якому сигнал відсилається лише у випадку, коли сумарний сигнал перевищує цей поріг.

Зазвичай нейрони організовані у шари, де різні шари виконують різні операції з вхідними даними. Сигнали проходять від першого (вхідного) шару до останнього (вихідного) шару, можливо, проходячи через шари кілька разів для подальшого перетворення інформації.

2.3.2 Тренування нейронних мереж

Нейронні мережі отримують навчання, опрацьовуючи приклади, кожен з яких включає відомий "вхід" та "результат". Це призводить до формування ймовірно зважених асоціацій між ними, які зберігаються в самій структурі даних мережі. Процес тренування полягає в порівнянні обробленого виходу мережі (часто називається передбаченням) і цільового виходу для визначення похибки. Ця похибка служить основою для коригування вагових асоціацій мережі за допомогою правила навчання.

Послідовні корекції призводять до того, що результати мережі стають все більше подібними до цільових. Після досягнення визначеної кількості корекцій тренування може бути завершено на підставі певного критерію, що є формою керованого навчання.

Ці системи "вчать" виконувати завдання, оглядаючи приклади, без явного програмування правил для конкретних сценаріїв. Наприклад, у випадку розпізнавання зображень вони можуть навчитися розпізнавати зображення котів, аналізуючи приклади зображень, які були вручну відзначені як "кіт" та "не кіт", і використовуючи цю інформацію для ідентифікації котів на інших зображеннях.

Це відбувається без наперед визначених правил про те, що таке кіт, такі як наявність хутра, хвоста, вус та котоподібного вигляду, а замість цього автоматично виникають ідентифікаційні характеристики з опрацьованих прикладів.

2.3.4 Кероване навчання

Кероване навчання, також відоме як контрольоване навчання, навчання під наглядом або навчання з учителем, є парадигмою машинного навчання, де модель навчається на основі вхідних об'єктів (наприклад, векторів змінних-передбачувачів) та відповідних бажаних виходів (також відомих як мітки, які надає людина – керівний сигнал).

Ці навчальні дані обробляються з метою створення функції, яка визначає залежність між новими вхідними даними та очікуваними виходами. Оптимальний результат передбачає, що алгоритм правильно визначає виходи для невідомих прикладів. Це вимагає від алгоритму навчання вміння узагальнювати з тренувальних даних на нові ситуації способом, який вважається "розумним" (див. індуктивне упередження).

Якість цього алгоритму оцінюється за допомогою поняття похибки узагальнення.



Рисунок 2.1 – Методи некерованого та керованого навчання

Щоби розв'язати задану задачу керованого навчання, слід виконати наступні кроки.

1. Визначити тип тренувальних прикладів. Перш ніж робити щось інше, користувач повинен вирішити, який тип даних використовувати як тренувальний набір. У випадку аналізу рукописного тексту, наприклад, це може бути один рукописний символ, ціле рукописне слово, ціле речення рукописного тексту або, можливо, повний абзац рукописного тексту.

2. Зібрати тренувальний набір (англ. training set). Він повинен бути репрезентативним для реального використання функції. Таким чином, збирають набір об'єктів входу, а також збирають відповідні дані виходу, або від людей-експертів, або з вимірювань.

3. Визначити подання ознак (англ. features) входу навчаної функції. Точність навчаної функції сильно залежить від того, як подано об'єкт входу. Як правило, вхідний об'єкт перетворюють на вектор ознак, що містить низку ознак, які описують об'єкт. Кількість ознак не повинна бути занадто великою через прокляття розмірності; але повинна містити достатньо інформації для точного передбачування виходу.

4. Визначити структуру навчаної функції та відповідний алгоритм навчання. Наприклад, інженер може обрати використання опорновекторних машин чи дерев рішень.

5. Завершити розробку. Виконати алгоритм навчання на зібраному тренувальному наборі. Деякі алгоритми керованого навчання вимагають від користувача визначення певних керівних параметрів. Ці параметри можна налаштувати, оптимізуючи продуктивність на підмножині тренувального набору (званій затверджувальним набором, англ. validation set), або за допомогою перехресного затвердження (англ. cross-validation).

6. Оцінити точність навчаної функції. Після налаштування параметрів та навчання продуктивність отриманої в результаті функції повинно бути виміряно на випробувальному наборі (англ. test set), окремому від тренувального.

2.4 Висновки за розділом 2

В результаті теоретичного обґрунтування обґрунтування сучасних методів режекції голосу з міксу показано, що існує багато різноманітних методів. Кожен з них може підійти під певні задачі різного рівня складності. Розглянуто роботу із саундизайном доріжок з нуля. Цей метод є найефективнішим оскільки дозволяє створювати звуковий супровід із нуля, що допомагає нам запобігти спотворенню вихідного сигналу.

Також було розглянуто роботу окремих плагінів типу Stereo Imager. Плагіни такого типу дозволяють нам працювати з мід та сайд частотами нашого сигнали. Вони також є дуже ефективними при видалення голосу із загального міксу, але додають певні спотворення у мід частотах.

Робота із штучними нейронними мережами є найіноваційнішими у роботі з методами режекції. Для використання штучних нейронних мережей їх спочатку треба навчити. Було розглянуто метод керованого навчання та кроки для його виконання

Проведено теоретичні оцінки ефективності кожного із сучасних методів та зроблено висновки.

Для уточнення отриманих результатів доцільно провести експериментальне дослідження порівняння роботи кожного з цих методів.

3 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ СУЧАСНИХ МЕТОДІВ РЕЖЕКЦІЇ ГОЛОСУ З МІКСУ

3.1 Дослідження методів режекції голосу

По черзі буде виконано режекцію голосу за допомогою кожного з сучасних існуючих методів. За допомогою спектрального аналізу буде виявлено спотворення. Данні будуть винесені в окрему таблицю для порівняння та виявлення найефективнішого методу режекції.

На рисунку 3.1 представлена спектрограма оригінальної звукової доріжки трейлеру.

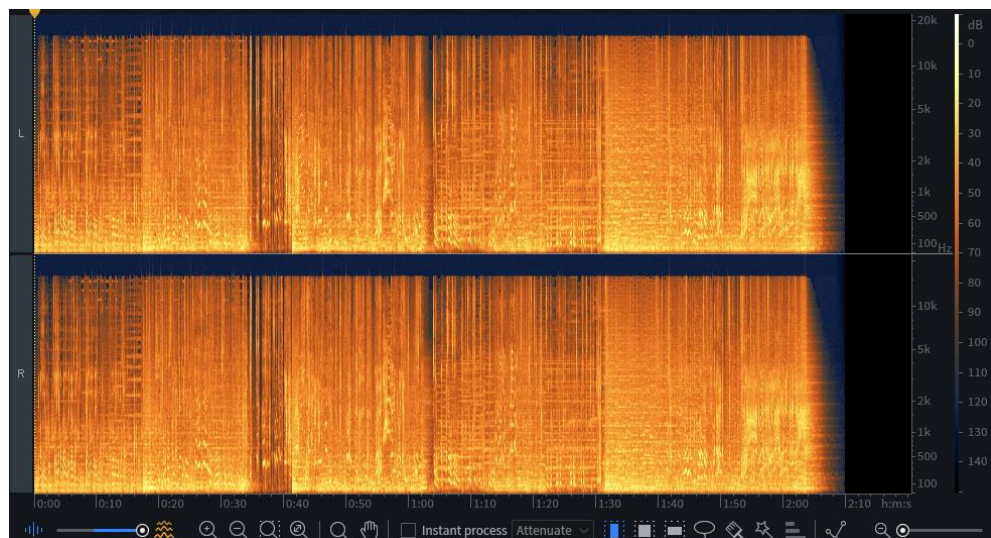


Рисунок 3.1 – Спектрограма оригінальної звукової доріжки

На спектрограмі аудіо-доріжки, створеної за допомогою саунд-дизайну зображеній на рисунку 3.2, видно цікаві артефакти, які відзначаються відмінностями в середньому та нижньому спектрі в порівнянні з оригінальною доріжкою. Ці відмінності створюють унікальний звуковий пейзаж, де відсутній голосовий сигнал, а також звуки оточення, що зазвичай супроводжують аудіо. Як наслідок, спектрограма відображає інші частоти та

звукові хвилі, які можуть бути пов'язані із саунд-ефектами або експериментальними аспектами саунд дизайну.

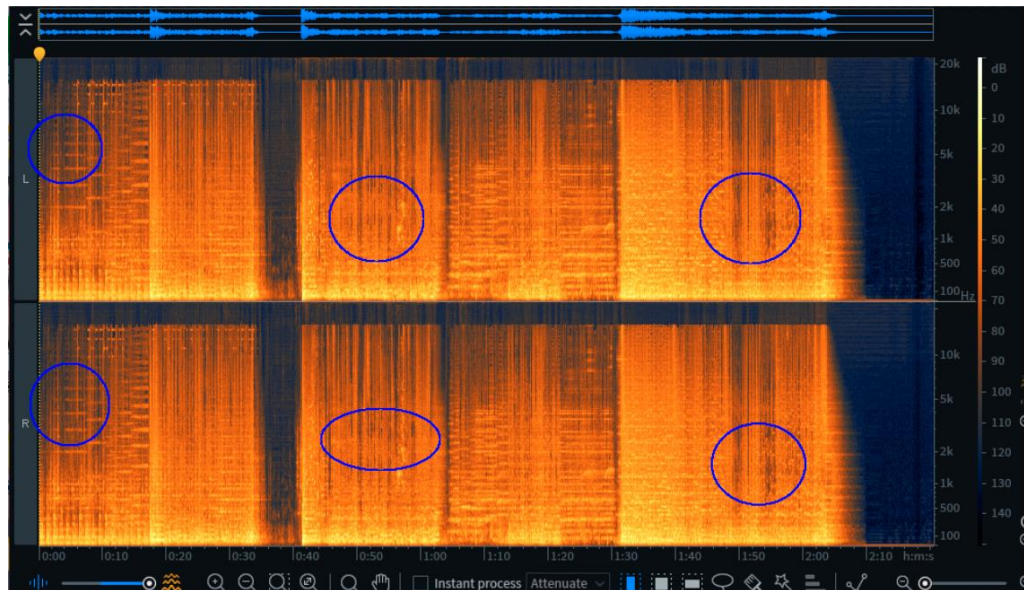


Рисунок 3.2 – Спектрограма доріжки створеної за допомогою саунд дизайну

Відсутність голосового сигналу вказує на те, що ця аудіо-доріжка, ймовірно, призначена для використання в якості фонового або атмосферного саунду, де фокус зміщений від мовлення на звукові аспекти, створюючи унікальний атмосферний ефект. Зокрема, вибір звуків оточення та саундтреку суттєво відрізняється від оригінального, створюючи новий аудіальний досвід для слухача.

Подібні техніки саунд-дизайну не лише розширюють можливості аудіовиробництва, але й додають глибину та індивідуальність до створених звукових композицій, надаючи їм унікальний естетичний відтінок та визначаючи їхнє місце в арсеналі звукового творчості.

На спектрограмі аудіо-доріжки, обробленої плагіном stereo imager показаної на рисунку 3.3, виокремлюються цікаві аспекти, які відображаються у вигляді характерної "ями" у середньому діапазоні, що відповідає за мовлення.

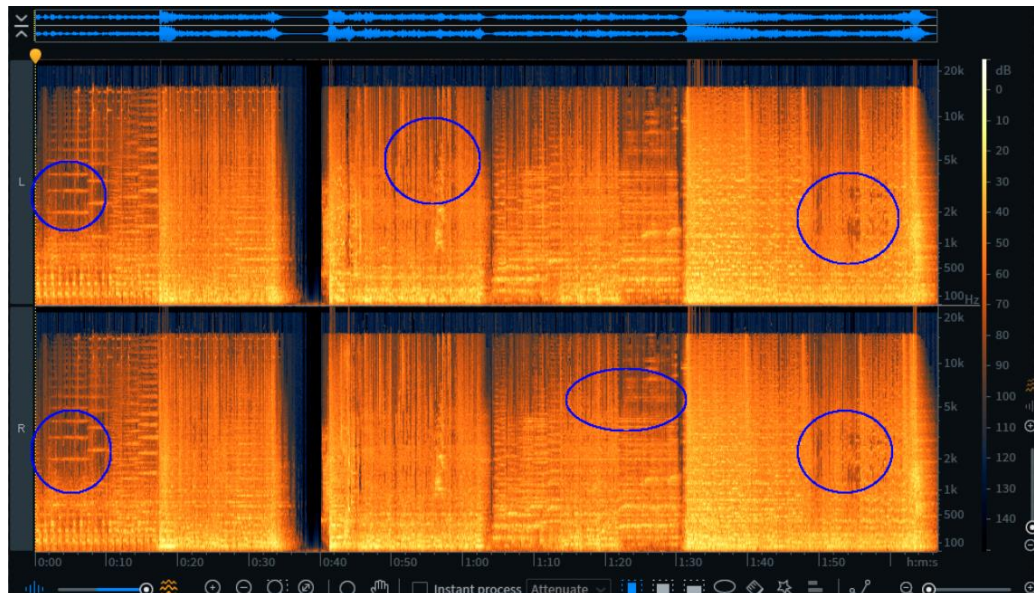


Рисунок 3.3 – Спектрограма доріжки обробленою Stereo Imager

Також помітні переривання у верхньому діапазоні частот, які зазвичай використовуються для голосу жіночих персонажів. Це пояснюється тим, що плагін stereo imager в основному впливає на середні частоти, викликаючи виникнення зазначених аномалій.

Застосування stereo imager, хоч і дозволяє створювати просторовий звук, супроводжується певними компромісами. Специфічна обробка середніх частот може призводити до втрати деяких елементів аудіосупроводу, включаючи важливі аспекти звукового оточення. Це може виявитися критичним у випадках, коли важливі аудіо-артефакти, такі як звуки оточення чи різноманітні компоненти саундтреку, відіграють ключову роль у створенні атмосфери та вираженні емоційної глибини сцени.

Детальний аналіз плагіну stereo imager розкриває його вплив на аудіо-ландшафт, де він, незважаючи на певні обмеження, може вносити цікаві просторові ефекти, але водночас вимагає обережного використання для уникнення небажаних артефактів та втрати важливих аудіоелементів.

На спектрограмі аудіо-доріжки, обробленої за допомогою штучного інтелекту (рис.3.4), виявляється цікава динаміка.

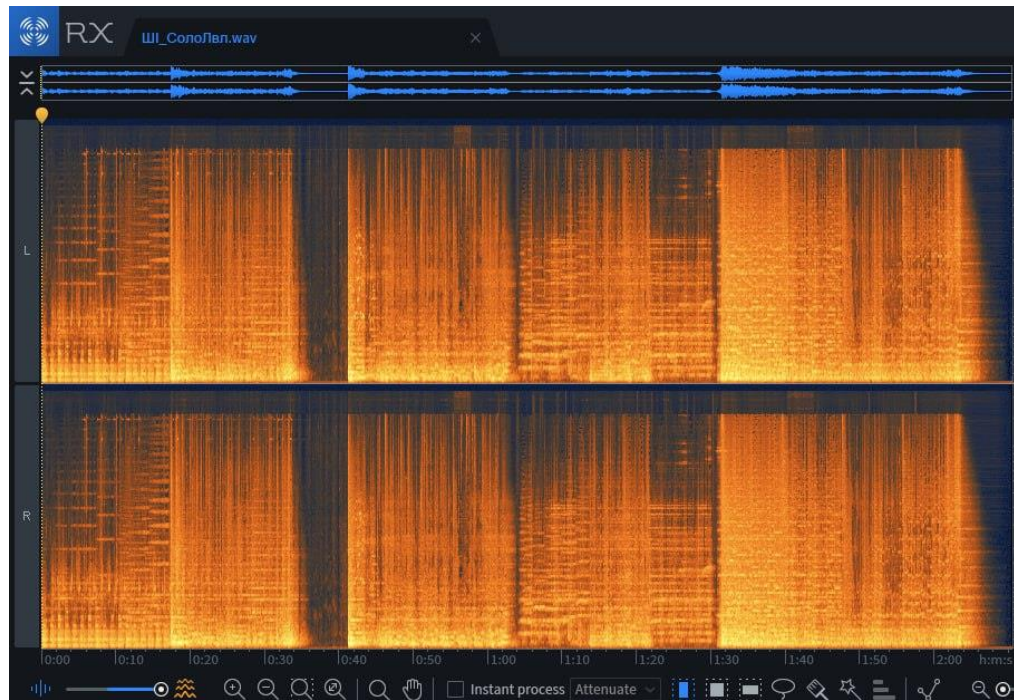


Рисунок 3.4 – Спектрограма доріжки обробленої за допомогою ШІ

Спостерігається мінімальна "яма" у середньому діапазоні, що може свідчити про деяку відмінність в обробці частот. Однак вражає велика яма у високому діапазоні, що може наводити на думку, що модель штучного інтелекту, ймовірно, була навчена переважно на чоловічому голосі, який зазвичай має вищий реєстр.

Таблиця 3.1 – Результати дослідження методів режекції

Метод режекції	Спотворення сигналу	Рисунок	Затрати часу	Загальна оцінка
Саунд дизайн	10	Рис.3.2	10	10
Stereo Imager	5	Рис.3.3	1	3
Нейронні Мережі	9	Рис.3.4	5	7

Цей підхід стає зрозумілим, оскільки чоловічі голоси вважаються менш вимагаючими для сприйняття різними моделями штучного інтелекту. Наразі він представляє собою один із найефективніших і найпростіших методів

режекції в обробці аудіо, але вимагає значних ресурсів у вигляді потужного обчислювального обладнання та об'ємної оперативної пам'яті для ефективної реалізації.

3.2 Висновки по розділу 3

Експериментально виявлено характерні особливості спотворень сигналу після режекції різними сучасними методами.

Отримані експериментальні оцінки приблизно відповідають теоретичним даним, отриманим в розділі 2.

Саунд-дизайн виявився найефективнішим методом, проте його використання вимагає значних трудовитрат. Цей метод, який включає в себе створення унікальних звукових ефектів та атмосфери, необхіден для досягнення високого рівня іммерсії в аудіо. Водночас він потребує обширної бази ефектів та можливостей запису для досягнення бажаного звучання.

Стосовно роботи із стереоімеджером, він визначається як найшвидший метод, який водночас не потребує додаткового обладнання чи великої кількості ресурсів. Але важливо враховувати, що цей підхід може додати найбільше спотворень у сигнал на виході, що потребує уважного балансу між ефективністю та якістю.

Нейронні мережі, з свого боку, займають проміжне положення з точки зору часу та якості. Хоча вони вимагають середню кількість часу для навчання та рендерингу, вони надають мінімальну кількість спотворень у порівнянні із стереоімеджером. Але для їх ефективної роботи необхідне потужне обчислювальне обладнання та велика кількість пам'яті для зберігання різноманітних нейронних моделей та матеріалу для їх навчання.

Дані висновки підтверджено на практиці, шляхом обробки оригінального аудіо сигналу трейлеру. Всі аудіоматеріали мали формат WAV, 44100 Гц, 24 біт.

ВИСНОВКИ

Дослідження методів режекції голосу з міксу є важливою складовою сучасного аудіовиробництва. Інноваційні підходи до цього процесу можуть значно покращити ефективність обробки аудіо та сприяти швидшому створенню та розповсюдженню більшого обсягу аудіо та відео контенту для різноманітної аудиторії.

Нові методи режекції голосу можуть призвести до оптимізації та прискорення робочого процесу, дозволяючи аудіовиробникам швидше впроваджувати креативні ідеї та експерименти у своїх проектах. Це, в свою чергу, відкриває можливості для широкого розповсюдження різноманітного контенту, що задовольняє потреби різних аудиторій та підвищує загальну доступність творчих та розважальних матеріалів.

Такі вдосконалення можуть сприяти розвитку та зростанню індустрії медіаконтенту, стимулюючи творчість та інновації в аудіовиробництві. В результаті цього, аудіо- та відеопродукція може стати більш доступною та різноманітною для глядачів та слухачів по всьому світу. Об'єкт дослідження – сучасні методи режекції звуку для дубляжу фільмів.

Мета роботи – показати перспективи обробки сигналу та використання нейронних мереж для видалення голосу.

В результаті теоретичного обґрунтування методів режекції було виявлено найефективніші серед існуючих.

Розглянуто різноманітні аспекти звукового тракту та його компонентів, серед яких важливе місце займають еквайзери та компресори. Еквайзери різних видів дозволяють точно налаштувати частотний баланс звукового сигналу, а компресори впливають на динаміку звучання, забезпечуючи контроль над рівнем гучності.

Серед методів ізоляції голосу, досліджених у роботі, важливе місце займають шумозаглушення, активне шумозаглушення та обробка частотної характеристики, яка змінюється у часі. Ці підходи дозволяють витіснити

непотрібний шум та забезпечувати кристально чистий звук голосу чи інших аудіоелементів.

Досліджено частотні спектри звукових сигналів, поділених на 8 рівнів: суббас, верхній бас, кабінетні частоти, середні частоти, середні високі, презенс, високі та айр. Це дозволяє зрозуміти, як різні частотні діапазони взаємодіють у комплексі та як їхні характеристики можуть впливати на загальний звуковий образ.

Також розглянуто амплітудно-частотні характеристики фільтрів, що розширює розуміння процесів обробки сигналу. Ці характеристики визначають, як фільтри реагують на різні частоти та впливають на їхню амплітуду, сприяючи досягненню конкретних звукових ефектів у музичному або аудіовиробничькому контексті.

Експериментально виявлено характерні особливості спотворень сигналу після режекції різними сучасними методами.

Отримані експериментальні оцінки приблизно відповідають теоретичним даним, отриманим в розділі 2.

Саунд-дизайн виявився найефективнішим методом, проте його використання вимагає значних трудовитрат. Цей метод, який включає в себе створення унікальних звукових ефектів та атмосфери, необхіден для досягнення високого рівня іммерсії в аудіо. Водночас він потребує обширної бази ефектів та можливостей запису для досягнення бажаного звучання.

Стосовно роботи із стереоімеджером, він визначається як найшвидший метод, який водночас не потребує додаткового обладнання чи великої кількості ресурсів. Але важливо враховувати, що цей підхід може додати найбільше спотворень у сигнал на виході, що потребує уважного балансу між ефективністю та якістю.

Нейронні мережі, з свого боку, займають проміжне положення з точки зору часу та якості. Хоча вони вимагають середню кількість часу для навчання та рендерингу, вони надають мінімальну кількість спотворень у порівнянні із стереоімеджером. Але для їх ефективної роботи необхідне

потужне обчислювальне обладнання та велика кількість пам'яті для зберігання різноманітних нейронних моделей та матеріалу для їх навчання.

Дані висновки підтверджено на практиці, шляхом обробки оригінального аудіо сигналу трейлеру. Всі аудіоматеріали мали формат WAV, 44100 Гц, 24 біт.

Отже, мета і задачі кваліфікаційної роботи виконані у повному обсязі.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. ANSI S1.1-1994 (ASA 111-1994) Revision of ANSI S1.1-1960 (R1976). Reaffirmed by ANSI March 25, 2004. Includes 2005 Erratum.
2. Jones M.R., Fay R.R., Popper A.N. Music Perception. New York: Springer, 2011. – 364 p.
3. Klingbeil M.K. Spectral Analysis, Editing, and Resynthesis: Methods and Applications. New York: Columbia University, 2009. – 468 p.
4. Clarkson MG, Clifton RK, Perris EE. Infant timbre perception: Discrimination of spectral envelopes. Perception & Psychophysics. 1988; 43(1). – p. 15-20.
5. V. Kartashov, V. Oleynikov, I. Koryttsev, S. Sheiko, O. Zubkov, S. Babkin, I. Selieznov. Use of Acoustic Signature for Detection, Recognition and Direction Finding of Small Unmanned Aerial Vehicles. 2020 IEEE 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). 2020. 4 p.
6. Kartashov V.M., Oleynikov V.N, Zubkov O.V., Koryttsev I.V., Babkin S. I., Sheiko S.A., Kolendovskaya M.M. Spatial-temporal Processing of acoustic Signals of Unmanned Aerial Vehicles/ Telecommunications and Radio Engineering. – New York. – 2020. – Vol. 79, №9. – P.769-780.
7. V. Kartashov, V. Oleynikov , I. Koryttsev, S. Sheiko, O. Zubkov, S. Babkin. Processing of Wide Band Acoustic Signals During Detection of Unmanned Aerial Vehicles // 2020 IEEE Ukrainian Microwave Week (UkrMW). Kharkiv, Ukraine, September 21 - 25, 2020. Volume 1 on 2020 IEEE 12th International Conference on Antenna Theory and Techniques (ICATT). pp. 35-39.
8. V.M. Kartashov, G.I. Sidorov, S.A. Sheiko, M.M. Kolendovskaya, O.Yu. Sergienko. Principles of construction and assessment of technical characteristics of multi-frequency atmospheric sodar in the humidity

- measurement mode. *Telecommunications and Radio Engineering*. Vol. 79. N.4. 2020. – pp. 323-333.
9. S. Sheiko. Study of the method for assessing atmospheric turbulence by the envelope of sodar signals // *Eastern-European Journal of Enterprise Technologies*. – 2/5 (92). – April, 2018. – p. 33–40.
 10. Сідоров Г.І., Шейко С.О., Шаповалов С.В., Полонська А.С., Дмитренко А.І. Акустичний метод вимірювання турбулентного стану атмосферного прикордонного шару // *Радиотехніка: Всеукр. міжвід. наук.-техн. зб.* 2018. – Вип. 192. – С. 46–50.
 11. Valerii V. Semenets, V. M. Kartashov, V. I. Leonidov. Registration of refraction phenomenon in the problem of acoustic sounding of atmosphere in airports zone. *Telecommunications and Radio Engineering*. Volume 77, Issue 5, 2018. – P. 461-468.
 12. Буйницький Д.В. Апаратно-програмний комплекс для дослідження акустичних пристроїв // 23-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті». – Харків, 16–18 квітня 2019 р. – с. 92-93.
 13. Чернов К.А. Дослідження акустичних студій звукозапис // 23-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті». – Харків, 16–18 квітня 2019 р. – с. 94-95.
 14. Ашихмін В.О. Адаптивні методи подавлення акустичного зворотного зв'язку: атестаційна робота на здобуття освітнього ступеня магістра. – Х.: ХНУРЕ. – 2019 р. – 69 с.
 15. Буйницький Д.В. Дослідження методів компенсації ревербераційних завад: атестаційна робота на здобуття освітнього ступеня магістра. – Х.: ХНУРЕ. – 2019 р. – 67 с.
 16. Шевченко П.Д. Система вимірювання та аналізу акустичних характеристик приміщень: атестаційна робота на здобуття освітнього ступеня магістра. – Х.: ХНУРЕ. – 2019 р. – 68 с.

17. Лісовий Р.О. Комплексне дослідження акустичних характеристик студійного приміщення: атестаційна робота на здобуття освітнього ступеня магістра. – Х.: ХНУРЕ. – 2019 р. – 81 с.
18. Древальський Р.В. Дослідження методу корекції звука для компенсації впливу приміщення /25-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті». Зб. матеріалів форуму. Т. 3. – Харків: ХНУРЕ. 2021. – с. 119 – 120.
19. Покровский Н.Б. Расчет и измерение разборчивости речи. – М., Связьиздат, 1962. – 390с.
20. Методичні вказівки з виконання атестаційної магістерської роботи за спеціальністю 8.05090102 «Апаратура радіозв'язку, радіомовлення і телебачення». Освітньо-кваліфікаційний рівень – магістр / Упоряд. В.М. Карташов, В.А. Тихонов, І.В. Савченко – Харків: ХНУРЕ, 2012. – 68 с.
21. Ovsinkiy V. The mixing engineering handbook, New York 2016, 309-556 p.