

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)

Методи дереверберації звукозаписів голосу.
(тема)

Виконав:
студент 2 курсу, групи МІМ-22-1
Шередега А.Г.
(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. Шаповалов С.В.
(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 Шередегі Артему Геннадійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Методи дереверберації звукозаписів голосу.

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 аналіз процесів реверберації в звукозаписі та їх впливу на якість сприйняття мови. Провести теоретичний аналіз методів усунення реверберації. Скласти класифікацію методів дереверберації, обрати методи, які найменше спотворюють якість корисного сигналу і можуть бути застосовані в медіаіндустрії. Дослідити обраний метод в системі MATLAB. Розроблено програму на основі обраного алгоритму для усунення пізньої реверберації. Провести оцінку точності усунення реверберації. Зробити висновки.

4. Перелік питань, що потрібно опрацювати в роботі _____

Вступ

1 Аналіз процесів реверберації в звукозаписі та їх впливу на якість сприйняття мови.

2 Методи дереверберації.

3 Метод спектрального віднімання і його різновиди.

4 Дослідження дереверберації звукового сигналу методом спектрального віднімання.

Висновки

Перелік посилань

Додатки


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

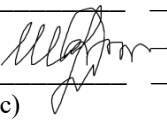
1. Постановка задачі (1 аркуш А4).
2. Аналіз процесів реверберації (1 аркуш А4).
3. Звукові плани (1 аркуш А4).
4. Якість і розбірливість мови (1 аркуш А4).
5. Класифікація методів дереверберації (1 аркуш А4).
6. Метод лінійного передбачення (1 аркуш А4).
7. Метод спектрального віднімання (1 аркуш А4).
8. Метод гомоморфної деконволюції (1 аркуш А4).
9. Метод Лембарта (1 аркуш А4).
10. Метод Ву та Ванга (1 аркуш А4).
11. Метод Фарая та Катаоки (1 аркуш А4).
12. Умови моделювання (1 аркуш А4).
13. Дослідження на гітарному акорді (1 аркуш А4).
14. Дослідження на мовному сигналі (1 аркуш А4).
15. Залежність RMS від порядку моделі ревербератора (1 аркуш А4).
16. Залежність RMS від звукопоглинання (1 аркуш А4).
17. Залежність RMS від об'єму приміщення (1 аркуш А4).
18. Висновки (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 дереверберації	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 _____  _____ Артем ШЕРЕДЕГА _____

Керівник роботи _____  _____ Сергій ШАПОВАЛОВ _____
(підпис)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 65 сторінок, 30 рисунків, 46 джерел.

ДЕРЕВЕРБЕРАЦІЯ РЕВЕРБЕРАЦІЯ, ГЕНЕРУВАННЯ ШТУЧНОЇ РЕВЕРБЕРАЦІЇ, МЕТОД СПЕКТРАЛЬНОГО ВІДНІМАННЯ, МЕТОДИ УСУНЕННЯ РЕВЕРБЕРАЦІЇ, МОДЕЛЮВАННЯ РЕВЕРБЕРАЦІЇ, ПІЗНЯ РЕВЕРБЕРАЦІЯ

Метою кваліфікаційної роботи є аналіз методів дереверберації звукозапису голосу з метою визначення методів, що найменше спотворюють якість корисного сигналу і можуть бути успішно застосовані в медіаіндустрії.

В роботі розглянуто методи усунення реверберації. Обрано методи, які найменше спотворюють якість корисного сигналу і можуть бути застосовані в медіаіндустрії. Досліджено метод на основі спектрального віднімання. Розроблено програму на основі алгоритму спектрального віднімання для усунення пізньої реверберації та проведено оцінку точності усунення реверберації. Досліджений метод усунення реверберації дозволяє значною мірою подолати проблеми, що виникають при загальноприйнятих класичних методів та алгоритмів розв'язання задачі усунення реверберації для звукових сигналів в медіаіндустрії.

ABSTRACT

Explanatory note to the qualification work: 65 pages, 30 figures, 46 sources.

DEREVERBERATION REVERBERATION, GENERATION OF ARTIFICIAL REVERBERATION, SPECTRAL SUBTRACTION METHOD, REVERBERATION ELIMINATION METHODS, REVERBERATION SIMULATION, LATE REVERBERATION

The purpose of the qualification work is to analyze the methods of reverberation of voice sound recording in order to determine the methods that least distort the quality of the useful signal and can be successfully applied in the media industry.

The work considers the methods of eliminating reverberation. The methods that least distort the quality of the useful signal and can be applied in the media industry are chosen. The method based on spectral subtraction was investigated. A program was developed based on the spectral subtraction algorithm for the elimination of late reverberation, and the accuracy of reverberation elimination was evaluated. The researched method of reverberation elimination allows you to largely overcome the problems arising from conventional classical methods and algorithms for solving the problem of reverberation elimination for sound signals in the media industry.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів.....	8
Вступ.....	9
1 АНАЛІЗ ПРОЦЕСІВ РЕВЕРБЕРАЦІЇ В ЗВУКОЗАПИСІ ТА ЇХ ВПЛИВУ НА ЯКІСТЬ СПРИЙНЯТТЯ МОВИ.....	11
1.1 Постановка задачі.....	11
1.2 Реверберація в замкнутих просторах.....	14
1.3 Вплив реверберації на сприйняття мови.....	20
1.4 Розбірливість і якість мовлення.....	21
1.5 Висновки по розділу 1.....	25
2 МЕТОДИ ДЕРЕВЕРБЕРАЦІЇ.....	27
2.1 Класифікація методів дереверберації.....	27
2.2 Метод явного моделювання мови.....	28
2.3 Метод покращення залишку лінійного передбачення.....	29
2.4 Метод спектрального віднімання.....	31
2.5 Методи просторової обробки.....	32
2.6 Гомоморфна деконволюція.....	33
2.7 Висновок по розділу 2.....	34
3 МЕТОД СПЕКТРАЛЬНОГО ВІДНІМАННЯ І ЙОГО РІЗНОВИДИ.....	35
3.1 Оцінка спектру реверберованого сигналу.....	35
3.2 Метод Лебарта, Бочера і Дебніха.....	38
3.3 Метод Ву та Ванга.....	41
3.4 Метод Блейкера та Георганті.....	43
3.5 Метод Фарая та Катаоки.....	44
3.6 Висновок по розділу 3.....	45
4 ДОСЛІДЖЕННЯ ДЕРЕВЕРБЕРАЦІЇ ЗВУКОВОГО СИГНАЛУ МЕТОДОМ СПЕКТРАЛЬНОГО ВІДНІМАННЯ.....	47
4.1 Умови експерименту.....	47

4.2	Моделювання гармонічного сигналу.....	48
4.3	Моделювання з використанням мовного сигналу.....	50
4.4	Аналіз роботи алгоритму дереверберації в залежності від вхідних параметрів.....	52
4.5	Висновок по розділу 4.....	55
	Висновки.....	56
	Перелік посилань.....	61
	ДОДАТКИ.....	66
	Додаток А. Лістинг програми <code>tawfAlgorithm.m</code> для усунення реверберації.....	67
	Додаток Б. Лістинг програми <code>delaynetwork.m</code> для генерації штучної реверберації.....	69
	Додаток В. Графічний матеріал.....	71
	Додаток Г. Відомість кваліфікаційної роботи.....	89

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,
СКОРОЧЕНЬ І ТЕРМІНІВ

AIR – Acoustic Impulse Response – імпульсна характеристика приміщення;

Alcons – articulation loss for consonants – втрати розбірливості;

ATF – Acoustic Transfer Function – акустична частотна функція передачі;

DRR – Direct to Reverberation Ratio – відношення прямого до відбитих звуків;

FFT Fast Fourier Transform – швидке перетворення Фур'є;

HERB – Harmonicity based dEReverBeration – деверберація на основі гармонійності;

HMM – Hidden Markov Model – прихована марківська модель;

IFFT – Inverse Fast Fourier Transform – зворотне FFT;

LP – Linear Prediction – лінійне передбачення

MFCC – Mel-frequency cepstral coefficient – мел-частотний кепстральний коефіцієнт;

PESQ – Perceptual Evaluation of Speech Quality – оцінка якості сприйняття мови;

PSD – Power Spectral Density – спектральна щільність потужності;

RMS – Root Mean Square – середньоквадратичний;

RT – Reverberation Time – час реверберації;

SNR – Signal to Noise Ratio – відношення сигнал-шум;

STFT – short-time Fourier transform – віконне перетворення Фур'є;

TIMIT – Acoustic-Phonetic Continuous Speech Corpus – мовна бази даних.

ВСТУП

Дереверберація – це процес, за допомогою якого ефекти реверберації видаляються зі звуку після того, як такий ревербераційний звук був уловлений мікрофонами. Дереверберація є підтемою акустичної цифрової обробки сигналів, яка найчастіше застосовується до промови, але також стосується деяких аспектів обробки музики.

Дереверберація звуку (мови або музики) є функцією, що відповідає сліпій деконволюції зображень, хоча використовувані методи зазвичай дуже різні. Сама реверберація викликається відбиттями звуку в кімнаті (або іншому замкнутому просторі) та кількісно визначається значенням часу реверберації у приміщенні та співвідношенням прямої/реверберації. Ефект дереверберації полягає у збільшенні відношення прямої/реверберації, щоб звук сприймався як ближчий та чіткіший.

Основне застосування дереверберації – це телефони з гучним зв'язком і настільні конференц-термінали, тому що в цих випадках мікрофони не знаходяться близько до джерела звуку – рота, що промовляє – але на відстані витягнутої руки або ще далі. Крім телекомунікацій, дереверберація широко застосовується в автоматичному розпізнаванні мови, оскільки розпізнавачі мови зазвичай схильні до помилок у сценаріях реверберації.

Дереверберація стала предметом наукових досліджень у період із 2000 по 2005 рр., хоча є кілька примітних ранніх статей. Перший науковий підручник з цієї теми був опублікований у 2010 році [1]. У 2014 році було проведено глобальне наукове дослідження, спонсороване IEEE Технічним комітетом з обробки звукових та акустичних сигналів.

Можна використовувати три різні підходи. для виконання дереверберації. У першому підході реверберація усувається за рахунок використання математичної моделі акустичної системи (або приміщення) та після оцінки параметрів акустичної моделі приміщення формування оцінки вихідного сигналу.

У другому підході реверберація пригнічується, розглядаючи її як тип (згорткового) шуму та виконуючи процес усунення шумів, спеціально адаптований для реверберації.

У третьому підході вихідний дереверберований сигнал безпосередньо оцінюється за сигналами мікрофона з використанням, наприклад, підходу машинного навчання глибокої нейронної мережі або альтернативно багатоканального лінійного фільтра.

Приклади найбільш ефективних сучасних методів включають підходи, що ґрунтуються на лінійному прогнозуванні.

Метою кваліфікаційної роботи є аналіз методів дереверберації звукозапису голосу з метою визначення методів, що найменше спотворюють якість корисного сигналу і можуть бути успішно застосовані в медіаіндустрії.

1 АНАЛІЗ ПРОЦЕСІВ РЕВЕРБЕРАЦІЇ В ЗВУКОЗАПИСІ ТА ЇХ ВПЛИВУ НА ЯКІСТЬ СПРИЙНЯТТЯ МОВИ

1.1 Постановка задачі

Сам процес звукозапису, якщо він виконується професійно, будується за схемою, щоби реверберації не виникало. Відомо, що на якість звукозапису найбільш впливає акустика приміщення. Професійні записи голосу відбуваються у спеціальній ізольованій від шуму і відбиттів акустичній кабінці з малою реверберацією (рис.1.1).



Рисунок 1.1 – Акустична кабінка для звукозапису

Дані умови далеко не завжди вдається виконати, тому у руки звукорежисера часто потрапляють фонограми, зроблені в непристосованих умовах (рис.1.2). Це можуть бути різного роду спонтанні інтерв'ю, виступи, лекції, зроблені без мікрофона-петлички. Або на самому інтерв'ю петличка може вийти з ладу, або сяде батарейка – і єдиним джерелом звуку залишиться доріжка, зроблена на накамерний мікрофон. В такому випадку реверберація буде дуже шкідливим ефектом, що значно знизить розбірливість мови і буде сильно дратувати слухачів.

Отже, задача дереверберації мови у медіавиробництві є дуже актуальною.

Вважаємо, що корисне джерело звуку розташоване на значній відстані від мікрофона (рис. 1.2). Корисне джерело створює звукові хвилі. Деякі з цих хвиль потрапляють безпосередньо на мікрофон. Отриманий прямий сигнал може бути погіршений реверберацією, фоновим шумом та іншими перешкодами.



Рисунок 1.2 – Постановка задачі

Для протидії спотворень сигналу, спричинених реверберацією, фоновим шумом тощо, потрібні високоефективні методи обробки акустичного сигналу.

Реверберація – це процес багатопроменевого поширення акустичних хвиль від джерела до одного чи кількох мікрофонів. Звук породжує механічну енергію, яка поширюється через речовину, наприклад, газ, у вигляді хвилі. Під впливом звукової хвилі відбувається зміна густини та тиску газу, які є функціями часу та положення. Різницю між миттєвим тиском $p(t)$ та статичним атмосферним тиском p_a називають звуковим тиском:

$$p_{зв}(t) = p(t) - p_a, \quad (1.1)$$

де $p_a = 101325$ Па в нормальних умовах.

Мікрофон використовується для перетворення тиску (або градієнта тиску), присутнього в повітрі безпосередньо перед мікрофоном в електричний

сигнал. Для простоти будемо вважати, що мікрофон ідеальний, тобто його електричний вихідний сигнал $s(t)$ ідентичний (за винятком безрозмірного масштабного коефіцієнта k) до локального звукового тиску [1]:

$$s(t) = k \cdot p(t), \quad (1.2)$$

де k – чутливість мікрофону, мВ/Па.

Для динамічних мікрофонів $k \sim 1 \dots 2$ мВ/Па, для конденсаторних мікрофонів $k \sim 10 \dots 20$ мВ/Па.

Структурна схема, яка описує застосування обробки акустичного сигналу, проілюстрована на рис. 1.3.

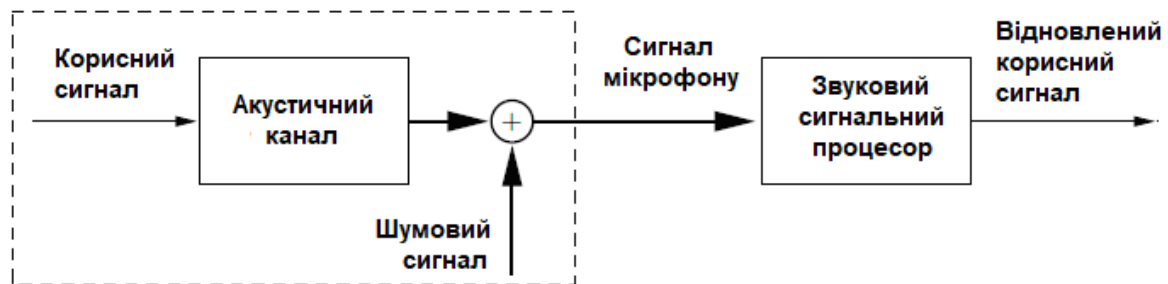


Рисунок 1.3 – Структурна схема, що описує застосування обробки акустичного сигналу

Тут звук, який створюється потрібним джерелом, позначеним як корисний сигнал або безеховий сигнал «передається» через акустичний канал та в комбінації із заважаючим сигналом це призводить до отриманого мікрофонного сигналу. Товсті лінії на рис. 1.3 позначають один або кілька сигналів, тоді як тонкі лінії позначають один сигнал.

Сигнали перешкод можуть описувати звуки перешкод або електричні перешкоди, наприклад шум датчика. Отриманий мікрофонний сигнал потім обробляється за допомогою процесору звукового сигналу для отримання оцінки потрібного сигналу.

Основна проблема в обробці акустичного сигналу виникає через спотворення корисного сигналу в акустичному каналі в замкнутому просторі.

Оскільки мікрофон не завжди можна розташувати поблизу потрібного джерела, прийняті мікрофонні сигнали зазвичай погіршуються через реверберацію при багатопроменевому поширенні корисного звуку до мікрофонів та шуму, внесеного джерелами перешкод.

При цьому найсучасніша обробка звукового сигналу має доступні алгоритми зменшення шуму та реверберації в обробленому сигналі. Основна відмінність між шумом і реверберацією полягає в тому, що спотворюючий компонент у разі реверберації залежить від корисного сигналу, тоді як у випадку шуму можна вважати, що він не залежить від корисного сигналу. Це

Реверберована мова може бути описана як віддалена мова з помітним забарвленням відлунням. Ефект спотворення відлунням зазвичай посилюється зі збільшенням відстані між джерелом і мікрофоном. Зі збільшенням часу надходження відбиттів у мікрофон, реверберація викликає розмиття фонем мови. Це серйозно погіршує розбірливість мови.

Алгоритми, які зменшують ці шкідливі ефекти, називаються деверберацією мови.

Щоб зменшити вплив реверберації за допомогою обробки звукового сигналу необхідно розуміти фізичні властивості процесу реверберації. Тому в розділі 1.2 розглядається реверберація в закритому просторі.

1.2 Реверберація в замкнутих просторах

Корисне джерело створює хвильові фронти, які поширюються ззовні джерела. Хвильові фронти відбиваються від стін кімнати, підлоги, стелі, предметів та накладаються біля мікрофону. На рис. 1.4 це проілюстровано на прикладі прямого шляху і єдиного відбиття. Через різницю в довжині шляхів розповсюдження до мікрофона та в кількості звукової енергії, поглиненої перешкодами кожний хвильовий фронт надходить на мікрофон з іншою амплітудою і фазою. Термін реверберація позначає наявність затриманих і ослаблених копій сигналу джерела отриманого сигналу.

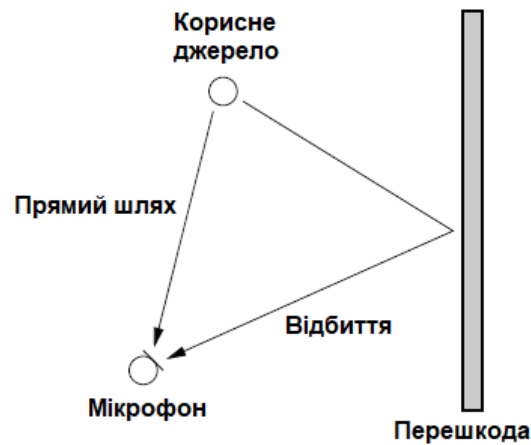


Рисунок 1.4 – Формування реверберованого сигналу

Реверберація – це процес багатопроменевого поширення акустичного сигналу від його джерела до мікрофона. Отриманий сигнал зазвичай складається з прямого звуку, відбиття, яке з’являється незабаром після прямого звуку (зазвичай називається рання реверберація), і відбиттів, які надходять після ранньої реверберації (зазвичай її називають пізньою реверберацією). Іноді поєднання прямого звуку і ранньої реверберації називають раннім звуковим компонентом. Розглянемо докладніше різні звукові компоненти.

Прямий звук. Перший звук, отриманий через вільний простір, тобто без відбиття, це прямий звук. У випадку, якщо джерело не знаходиться в зоні прямої видимості спостерігача прямого звуку немає. Затримка τ між початковим збудженням джерела і його спостереженням залежить від відстані r та швидкості звуку c [1]:

$$\tau = \frac{r}{c}, \quad (1.3)$$

де $c = 340$ м/с в нормальних умовах.

Рання реверберація. Трохи пізніше приходять звуки, які відбилися від однієї або декількох поверхонь (стіни, підлога, меблі тощо). Ці відбиті звуки відокремлені як за часом, так і за напрямком від прямого звуку.

Відбиті звуки утворюють звуковий компонент, який зазвичай називають ранньою реверберацією. Структура ранньої реверберації змінюватиметься залежно від зміни положення джерела або мікрофона у просторі та дає нам інформацію про розмір простору та розташування джерела в просторі.

Рання реверберація не сприймається як відокремлений звук від прямого, якщо затримка відбиттів не перевищує значення приблизно 80-100 мс відносно часу прибуття прямого звуку. Це значення обумовлене інерційністю людського слуху.

Рання реверберація фактично сприймається як підсилення прямого звуку і тому вважається корисною стосовно розбірливості мови. Це часто називають ефектом пріоритету, що полегшує ведення розмов у закритих приміщеннях, ніж на вулиці.

Рання реверберація головним чином важлива в так званій акустиці невеликих приміщень, де стіни, стеля та підлога дуже близькі. Рання реверберація також викликає спектральне спотворення, яке називається забарвленням.

Пізня реверберація. Пізня реверберація є результатом відбиттів, які надходять з більшими затримками після надходження прямого звуку. Вони сприймаються або як окреме відлуння або як реверберація та погіршують розбірливість мови.

Акустичний канал між джерелом і мікрофоном можна описати за допомогою відгуку на акустичний імпульс [1] (Acoustic Impulse Response – AIR). Це сигнал, який вимірюється мікрофоном у відповідь на джерело, яке створює «звуковий імпульс». Процес реверберації може бути розділений на три сегменти: прямий звук, ранні відбиття та пізні відбиття, як показано на рис.1.5.

Результуючий сигнал є згортою цих сегментів із корисним сигналом. З точки зору обробки сигналу, ранні відбиття виглядають як окремі затримані

імпульси у відгуку, тоді як пізні відбиття виглядають як безперервний ревербераційний хвіст.

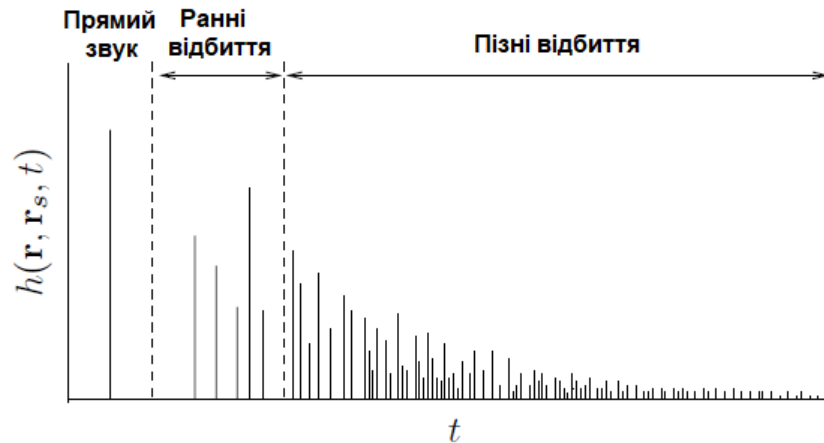


Рисунок 1.5 – Схематичне зображення реакції на акустичний імпульс

Крім того, це важливо зауважити, що енергія відбиттів спадає з експоненціальною швидкістю. Ця експоненціальна залежність є добре відомою властивістю імпульсного відгуку, яка призвела до появи поняття часу реверберації.

Час реверберації кількісно визначає інтенсивність реверберації в кімнаті та позначається RT_{60} . Він визначається як час, необхідний для згасання звукової енергії на 60 дБ після вимкнення джерела звуку (рис.1.6).

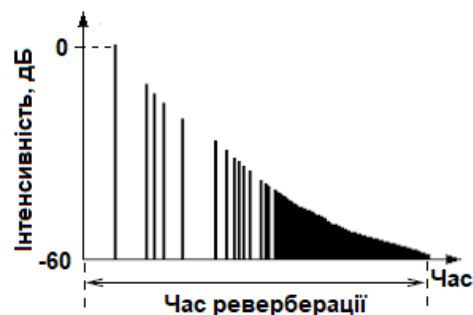


Рисунок 1.6 – Визначення часу реверберації RT_{60}

Часово-просторовий варіант імпульсного відгуку AIR $h(r, r_s, t, t_0)$ визначається як відповідь акустичного каналу між джерелом у положенні $r_s(x_s, y_s, z_s)$ та мікрофоном у положенні $r(x, y, z)$ у момент часу t через

одиничний імпульс, прикладений у момент часу t_0 . Спостережуваний сигнал у положенні r у момент часу t визначається як [1]

$$s(r, t) = \int_{-\infty}^{\infty} \int_{V_s} h(r, r_s, t') s_s(r_s, t') dr_s dt', \quad (1.4)$$

де $s_s(r_s, t')$ – акустичний сигнал у позиції r_s , час t' , а

V_s позначає обсяг приміщення.

Перетворення Фур'є від імпульсної характеристики AIR в момент часу t називається функцією передачі звуку (Acoustic Transfer Function – ATF) і позначається $H(r, r_s, t, f)$ [1]:

$$h(r, r_s, t, t') \xrightarrow{F} H(r, r_s, t, f), \quad (1.5)$$

де f позначає кутову частоту.

Функція ATF визначає частотну характеристику системи, що пов'язує джерело звуку зі звуковим тиском у мікрофоні, і це, ймовірно, найчастіше використовувана функція для опису акустичного каналу.

Для ревербераційних середовищ ATF є випадковою функцією, яку неможливо передбачити заздалегідь без детального знання акустичних та геометричних параметрів приміщення. Ступінь випадковості може характеризуватися спектральним відхиленням.

Незважаючи на те, що ревербераційне звукове поле має випадковість, воно також має базову структуру. Розібратися в структурі ATF можна за допомогою рівняння акустичної хвилі, яке описує поширенням звукової хвилі через матеріальне середовище. Існують різні акустичні моделі для приміщень. Більшість із цих моделей, залежать від сотень, а іноді й тисяч параметрів.

У більшості випадків неможливо виміряти АТФ під час роботи алгоритму обробки звукового сигналу, оскільки корисний сигнал невідомий на стороні приймача. У деяких випадках корисний сигнал відомий на стороні приймача, тоді АТФ можна розрахувати.

Інша проблема полягає в тому, що АТФ швидко змінюється під час переміщення джерела на значну відстань від початкової точки вимірювання [2-4]. Крім положення джерела АТФ чутлива до температури, розміщення меблів у кімнаті та інших рухів в кімнаті. Таким чином, навіть якщо можливо вимірювати АТФ в режимі реального часу, часто може знадобитися повторне вимірювання.

Якщо відстань між джерелом і мікрофоном змінюється, змінюється енергія прямого сигналу, тоді як комбінована енергія ранніх і пізніх відбиттів приблизно постійні. Відстань, на якій енергія прямого сигналу дорівнює сумарній енергії раннього і пізнього відбиттів, називається критичною відстанню.

Слід розуміти, що якщо відстань між джерелом і мікрофоном більша за критичну відстань, то енергія відбитих сигналів більша ніж енергія прямого сигналу. Для оцінки ступеня реверберації в звукозаписі вволять поняття акустичного відношення :

$$A = \frac{E_{ран} + E_{пізн}}{E_{пр}}, \quad (1.6)$$

де $E_{ран}$, $E_{пізн}$ – енергія ранніх і пізніх відбиттів відповідно;

$E_{пр}$ – енергія прямої звукової хвилі.

Оптимальне значення акустичного відношення передачі мовлення перебувають у межах 0,5...4, а музичних передач 2...8. Якщо акустичне відношення нижче цієї межі, то передача звучить уривчасто, сухо. Якщо вона

більша за верхню межу, то мова стає погано розбірливою, а музична передача – «забрудненою».

1.3 Вплив реверберації на сприйняття мови

Реверберуючу мову можна описати як віддалене звучання з помітним відлунням і забарвленням. Вплив реверберації на мову чітко чутний і видимий на спектрограмі та формі хвилі мовного сигналу. На рис. 1.7 (а) зображено спектрограма і форму хвилі (включаючи розшифровку) безехового мовного сигналу. Мовний сигнал був взятий з мовної бази даних ТІМІТ [5].

Мовленнєві форманти, які визначаються як резонансні частоти, пов'язані з голосовим трактом [6], можна чітко визначити на спектрограмі. Також видно, що фонемі добре розділені за часом.

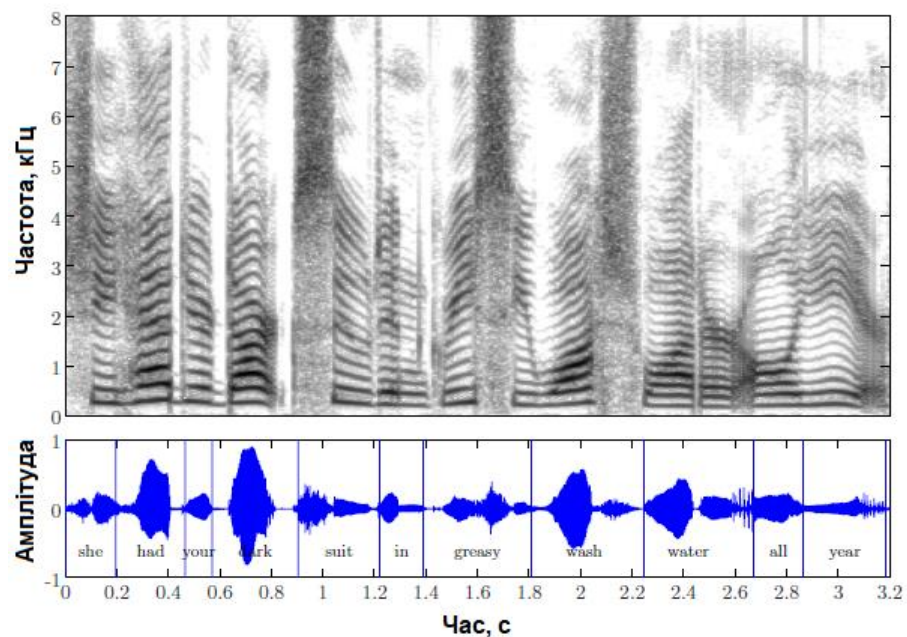


Рисунок 1.7 – Спектрограма і хвильоформа безехового сигналу

Безеховий сигнал на рис. 1.7 був переданий в офісну кімнату та її відгук вимірювали на відстані 0,5 м від джерела. Спектрограма і форма хвилі отриманого сигналу показана на рис. 1.8.

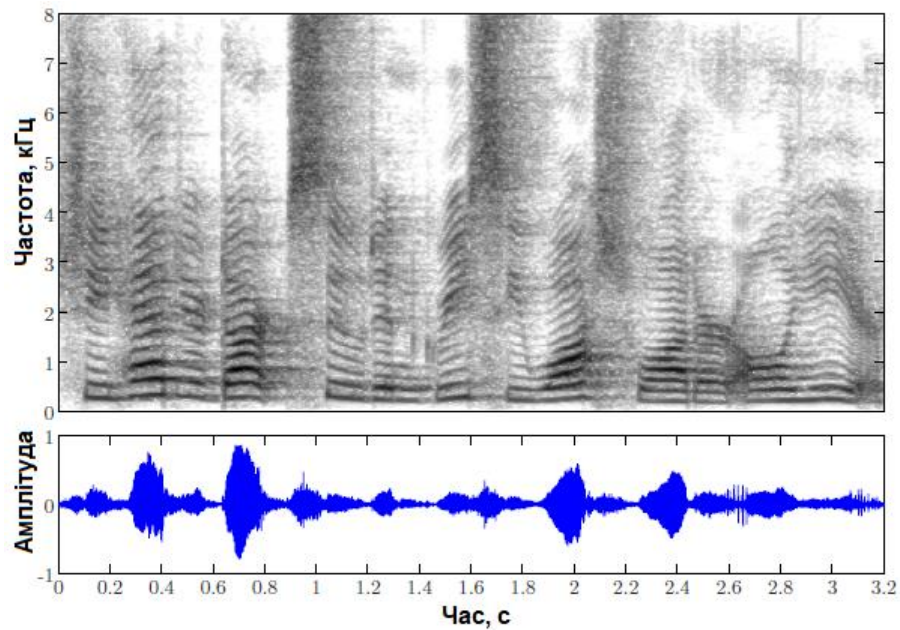


Рисунок 1.8 – Спектрограма і хвильоформа сигналу з відлунням

Спотворення мовного сигналу, викликане акустичним каналом, чітко видно. Розмитість формант мови видно на спектрограмі, а розмитість фонем у часі видно як на спектрограмі, так і на хвилі. За рахунок цього розмазування порожні проміжки між словами та складами заповнюються реверберацією та перекриваються з наступними фонемами. Ці спотворення призводять до чутної різниці між безеховою і лунаючою мовою, погіршується розбірливість і чіткість мови.

Для розробки ефективних алгоритмів деверберації має велике значення добре розуміти вплив реверберації на сприйняття мови. Тому розглянемо, який вплив має замкнутий простір на розбірливість і якість мовлення.

1.4 Розбірливість і якість мовлення

Фізичні властивості замкнутого простору, а також розташування джерела і слухача (мікрофону) у просторі мають великий вплив на реверберацію [8]. Реверберація і фоновий шум спричиняють помітні зміни в якості мови та визначають розбірливість мовлення в замкнутому просторі.

Зручно було б припустити що реверберація лише знижує розбірливість, але це припущення є невірним [9].

Інженери-акустики часто вважають відбиття бажаними, оскільки вони збільшують амплітуду сигналу, що досягає слухача. Це збільшення амплітуди може збільшити розбірливість мови, якщо вона підвищує рівень мови вище рівня навколишнього шуму [10]. Робота Лохнера та Бюргерса [11] демонструє цей ефект для одиничних відбиттів тощо. В роботі [12] продемонстровано цей ефект для складних ранніх відбиттів.

Ранні відбиття не покращують розбірливості, коли рівень звукового тиску в безеховому і реверберуючому записі однакові. Властивість інтеграції, яку зазвичай називають інертністю, зводиться у слуховій системі людини до інтеграції ранньої реверберації та прямого звуку, а також збільшує відчутну силу прямого звуку. Цю властивість відкрили у 1935 році Айгнер і Стратт [13]. Вони були першими, хто запропонував вимірювання на основі співвідношення акустичних енергій кількісно оцінити вплив фонового шуму та акустики приміщення на розбірливість мови.

Свою міру вони назвали відношенням Q , яке описується виразом

$$Q = \frac{E_d + E_e}{E_l + E_n}, \quad (1.7)$$

де E_d – пряма звукова енергія,

E_e – енергія хвиль ранніх відбиттів, які у цьому випадку визначаються як надходження до вуха не пізніше ніж за 60 мс після прямого звуку,

E_l – це пізня частина відбитої звукової енергії, яка надходить пізніше 60 мс,

E_n – енергія шуму.

Автори роботи [13] поставили нижній поріг $Q=1$ для задовільного звукового враження. Рівняння (1.7) по суті формує основу більшості показників розбірливості мови, які були розроблені пізніше.

Важливо відзначити, що звукове враження можна покращити, зменшивши або E_l , або E_n .

Приголосні відіграють набагато більшу роль у розбірливості мови, ніж голосні. Якщо чітко чути приголосні, легше зрозуміти мову. У 1971 році Reutz [14] запропонував вимірювання, яке називається втратою артикуляції приголосних (Alcons). Воно кількісно визначає зниження сприйняття приголосних через реверберацію.

Розрахунок даної міри залежить від відстані між джерелом і мікрофоном. Втрату артикуляції можна зменшити, тобто підвищити розбірливість мови, зменшуючи відстань джерело-мікрофон або час реверберації, а також збільшуючи об'єм приміщення.

У 1980 році Берклі досліджував сприйняття мови на основі застосування програма моделювання кімнати, що забезпечує набір добре контрольованих реалістичних відгуків кімнати з різними акустичними параметрами приміщення [15]. Використання результатів, отриманих під час аудіювання (тести Берклі) дійшли висновку, що сприйняття реверберації в основному базується на двох показниках.

Двома компонентами є фарбування звуку і відлуння. Берклі показав, що спектральне відхилення добре корелює з суб'єктивним сприйняття цього компонента як фарбування звуку. Ехо-компонент безпосередньо пов'язаний з часом реверберації RT_{60} . Звернемо увагу, що кількість пізньої реверберації збільшується при підвищенні RT_{60} .

У 1982 році Аллен [16] запропонував формулу для прогнозування суб'єктивного впливу реверберації на мову. Цей вплив описується рівнянням

$$P = P_{\max} - \sigma RT_{60}, \quad (1.8)$$

де P – суб'єктивна оцінка в деяких довільних одиницях,

P – максимум суб'єктивної оцінки.

Відповідно до цієї формули, якість сприйняття мови зменшує або спектральне відхилення σ , або час реверберації RT60.

В роботі [17] показано, що спектральне відхилення пов'язане з коефіцієнтом DRR (Direct to Reverberation Ratio), який визначається як енергія прямого звуку (E_d), поділену на загальну енергію відбитого звуку ($E_e + E_l$):

$$DRR = \frac{E_d}{E_e + E_l}. \quad (1.9)$$

Слід зазначити, що DRR в межах однієї кімнати є приблизно постійним і досягає свого максимального асимптотичного значення, якщо відстань джерело-мікрофон більша за критичну відстань, яка визначається як відстань, на якій енергія прямого звуку дорівнює повній енергії відбиття.

Коли відстань джерело-мікрофон менша за критичну відстань, спектральне відхилення можна використовувати для визначення DRR. У цій же кімнаті коротша відстань джерело-мікрофон призводять до більш високого DRR і меншого спектрального відхилення i , отже, забарвлення у випадку відстань джерело-мікрофон менше за критичну відстань.

З наведеного вище обговорення можна зробити висновок, що пізня реверберація та шум є основними основні причини погіршення розбірливості мовлення. Крім того, перцептивна якість мовлення, яка пов'язана із суб'єктивними перевагами та звуковим враженням, є пов'язаними з двома фізичними властивостями реверберації, тобто забарвленням і часом реверберації.

Слід зазначити, що ці властивості не є незалежними один від одного. Спектральне фарбування залежить від часу реверберації, гучності кімнати та відстані від мікрофона. Час реверберації RT60 важливий не тільки з точки зору сприйняття, але він також характеризує «форму» імпульсної характеристики AIR. Таким чином, час реверберації RT60 є важливим показником, який відіграє вирішальну роль у даній кваліфікаційній роботі.

1.5 Висновки по розділу 1

1. Процес звукозапису найчастіше будується за схемою, щоби реверберації не виникало. Професійні записи голосу відбуваються у спеціальній ізольованій від шуму і відбиттів акустичній кабіні з малою реверберацією.

2. Дані умови виконуються не завжди вдається виконати, і у руки звукорежисера часто потрапляють фонограми, зроблені в непристосованих умовах. Це можуть бути спонтанні інтерв'ю, виступи, лекції, зроблені без мікрофона-петлички. Або на самому інтерв'ю петличка може вийти з ладу, або сяде батарейка – і єдиним джерелом звуку залишиться доріжка, зроблена на накамерний мікрофон. В такому випадку реверберація буде дуже шкідливим ефектом, що значно знизить розбірливість мови і буде сильно дратувати слухачів.

3. Задача дереверберації мови у медіавиробництві є дуже актуальною. Метою кваліфікаційної роботи є аналіз методів дереверберації звукозапису голосу з метою визначення методів, що найменше спотворюють якість корисного сигналу і можуть бути успішно застосовані в медіаіндустрії.

4. В першому розділі проведено аналіз процесів реверберації в звукозаписі та їх впливу на якість сприйняття мови. Процес реверберації можна характеризувати імпульсним відгуком приміщення. Він складається з прямого звуку, ранніх відбиттів і пізніх відбиттів. Співвідношення між цими складовими залежить як від акустичних властивостей приміщення, так і від відстані від джерела до мікрофону.

5. Виділяють поняття критичної відстані, на якій енергія відбитих хвиль дорівнює енергії прямої хвилі. Є також поняття акустичного відношення, це відношення енергії відбитих хвиль до енергії прямої хвилі. Відповідно виділяють три звукових плани – близький, середній і віддалений.

6. Як показують дослідження, на якість сприйняття мови негативно впливають тільки пізні відбиття. Ранні відбиття зливаються з прямим звуком,

підсилюють його, виділяють на фоні шумів і фарбують спектр, накладаючи інформацію про приміщення. Тому показник якості в чисельнику містить прямий звук і ранні відбиття, а в знаменнику – те, що погіршує звук – пізні відбиття і шум. Отже, пізня реверберація та шум є основними основні причини погіршення розбірливості мовлення.

7. Кваліфікаційна робота виконується на кафедрі МІРЕС ХНУРЕ. На кафедрі проводяться дослідження в таких наукових областях, як виявлення та розпізнавання БПЛА за результатами акустичного спостереження [18-21]. Цілий ряд студентських доповідей [22-25] і атестаційних робіт магістрів минулих років [26-31] присвячені дослідженню систем звукозапису. Отже, дослідження в даній роботі пов'язані і ґрунтуються на традиційному напрямку робіт колективу і студентів кафедри МІРЕС.

2 МЕТОДИ ДЕРЕВЕРБЕРАЦІЇ

2.1 Класифікація методів дереверберерації

Методи дереверберерації можуть бути класифіковані в різний спосіб. Методи дереверберерації можна розділити на одно-мікрофонні і мульти-мікрофонні методи. Оскільки в нашому випадку йде мова про дереверберерацію в звукозаписі, тобто після процесу, будемо розглядати одноканальні системи. Більшість алгоритмів на декількох мікрофонах використовують просторову обробку сигналів і можуть бути використані тільки в процесі запису.

З іншої точки зору, способи дереверберерації можна розділити на ті, які в першу чергу зосереджені на забарвленні мови (рання реверберерація) і ті, які зосереджені на усуненні пізньої реверберерації.

В одному із способів класифікації [32], методи дереверберерації розглядаються незалежно від оцінки акустичної імпульсної характеристики приміщення. Цей критерій є результатом двох основних категорій, які називаються:

- придушення реверберерації;
- дереверберерація.

Методи першої категорії не оцінюють імпульсну характеристику приміщення (AIR), в той час як у другій категорії вони повинні оцінити AIR для того, щоб усунути реверберерацію в сигналі.

Найбільша увага буде приділена алгоритмам дереверберерації з одним мікрофоном, оскільки це найбільш поширений варіант у медіавиробництві. Також важливим аспектом є усунення ефекту реверберерації в звукозаписах мови.

Як сказано раніше, методи усунення реверберерації, які не використовують оцінки AIR, класифікуються як методи придушення реверберерації. Ці методи в свою чергу розділені на підкатегорії з урахуванням

кількості знань про джерело сигналу або канал поширення сигналу, і з відмінністю в методах обробки сигналів, які беруть у цьому участь.

На рис.2.1 наведена класифікація методів дереверберації.

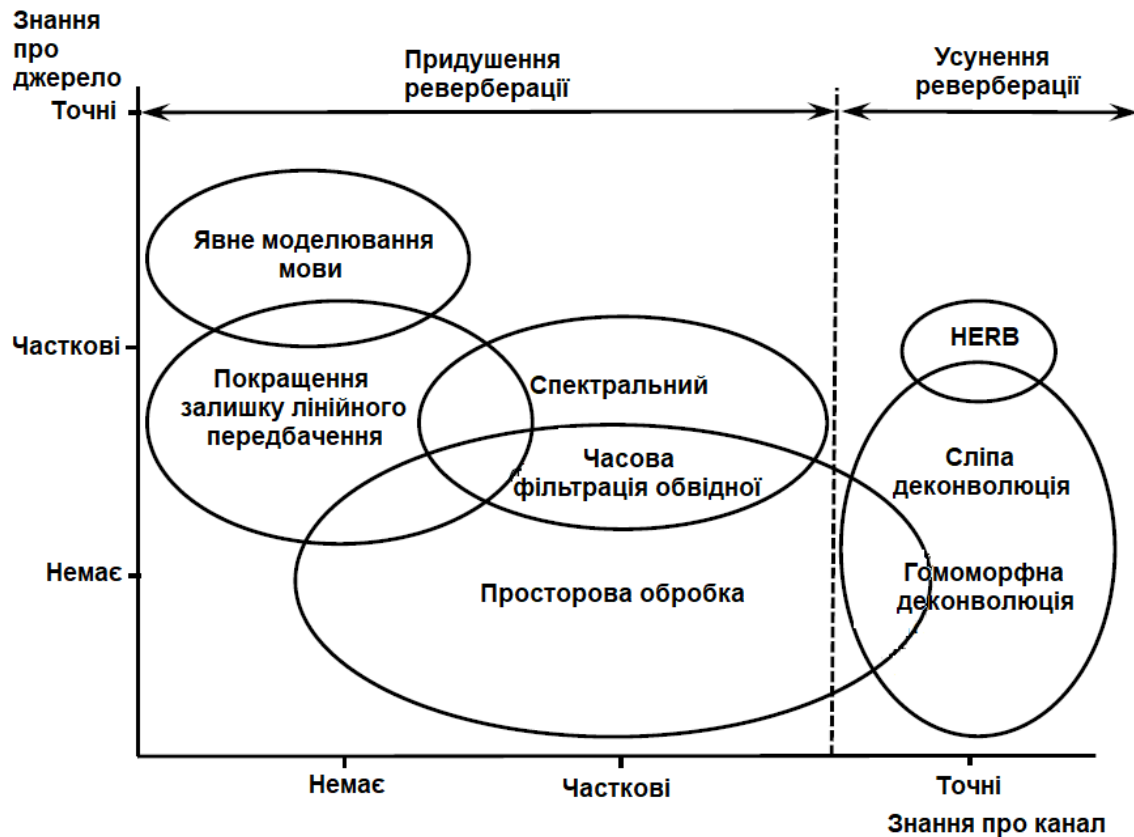


Рисунок 2.1 – Класифікація методів дереверберації з врахуванням кількості апріорних знань про канал поширення сигналу і джерело сигналу

2.2 Метод явного моделювання мови

Деякі методи усунення реверберації засновані на моделюванні мовного сигналу за допомогою базової структури безехового мовного сигналу. Дуальне збудження моделі мови була запропонована Хардвіком [33], ця модель була використана для поліпшення мови. При додаванні ефекту зміни висоти звуку в цій моделі, це доповнює узагальнену дуальну модель мови збудження. Примітно, що обидві з зазначених вище моделей базуються тільки на озвученні мовних сегментів.

Було запропоновано єдиний імовірнісний підхід для зниження рівня шуму і усунення реверберації для мовних сигналів з використанням

байєсівського оптимального оцінювання сигналу для шумозаглушення також для усунення негативного ефекту реверберації. Основна ідея цього методу полягає в попередньому навчанні моделі мови на великому наборі даних з безеховою мовою. Ця структура застосовується для одного так і для багатьох мікрофонів, проте усунення реверберації дає однаково добрі результати [34].

У той час як експерименти показують, що оптимальне байєсівське оцінювання може перевершити стандартні методи, такі як метод спектрального віднімання, в умовах придушення шуму, на жаль, продуктивність усунення реверберації не може оцінюватися окремо. Крім того, недолік цього методу полягає в тому, що вона в значній мірі залежить від професійної підготовки моделі.

2.3 Метод покращення залишку лінійного передбачення

Моделювання мови у вигляді згенерованої послідовності, яка формується змінним в часі полюсним фільтром – це поширений спосіб опису мовного сигналу [35]. Моделі збудження використовують послідовності невокалізованої мови на базі випадкової послідовності шуму і дзвінких мовних квазіперіодичних імпульсів. Фільтр, який використовується потім для формування мовного сигналу, являє собою модель людського голосового тракту. На рис. 2.2 показана ця мовна модель.

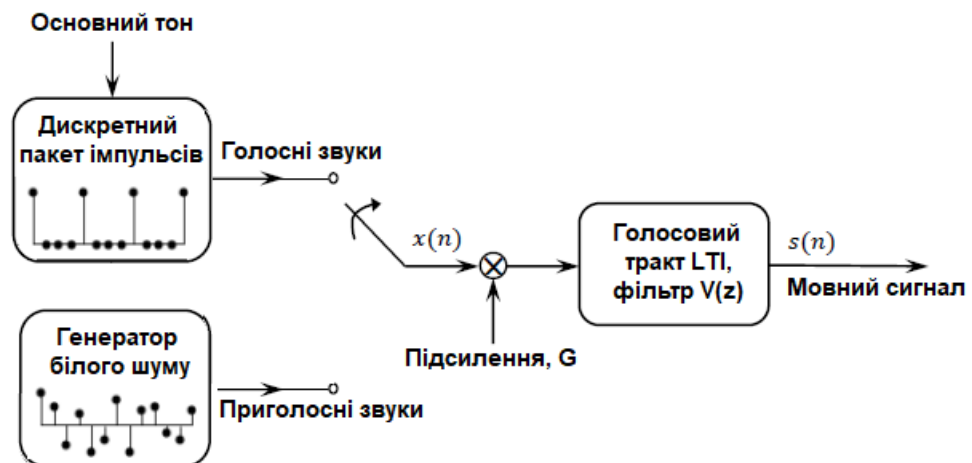


Рисунок 2.2 – Модель генерації мови з використанням методу явного моделювання

Голосовий тракт моделюється полюсовим фільтром, коефіцієнти якого оцінюються за допомогою аналізу лінійного передбачення (LP) записаної мови і називаються коефіцієнтами лінійного передбачення (LPC). У цій моделі залишок лінійного передбачення являє собою послідовність генерованих імпульсів, які можуть бути отримані шляхом зворотної-фільтрації мовного сигналу.

Обґрунтування використання техніки зворотної фільтрації базується на тому спостереженні, що в ревербераційному середовищі, залишок лінійного передбачення дзвінких мовних сегментів містить вихідні імпульси в додаток до кількох інших піків, отриманих за допомогою багатопроменевого відбиття. Важливим припущенням в цьому методі є те, що LPCS не впливає на реверберацію.

Таким чином, в методах цього класу усунення реверберації реалізується шляхом придушення цих піків в послідовності збудження (залишку лінійного передбачення), які відбуваються через багатопроменеве поширення відбиттів і синтезу мови за допомогою модифікованого залишку лінійного передбачення і змінного в часі всеполюсного фільтру (LP-фільтру) з коефіцієнтами (LPC), розрахованими з реверберованої мови.

При цьому, $x(n)$ представляє зразки ревербераційного сигналу, записаного з M мікрофонів в дискретному часі n . Блок аналізу LPC виконує функцію оцінки полюсів фільтра і змінює їх в часі і видає сигнал про помилку передбачення $\tilde{e}(n)$ (рис.2.3).

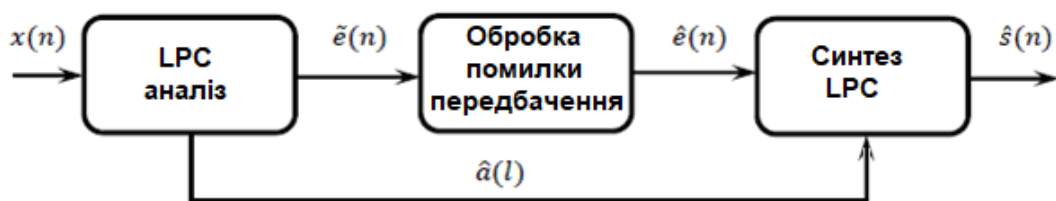


Рисунок 2.3. Загальна структура метода усунення реверберації що базується на покращенні залишку лінійного передбачення

Далі, ґрунтуючись на деяких критеріях і ознаках залежно від алгоритму, помилки передбачення змінюються і оцінюється чистий сигнал помилки передбачення $e(n)$. На наступному етапі, синтезується мовний сигнал з використанням розрахованих полюсів $\hat{a}(n)$ і оціненого чистого залишку передбачення $e(n)$ [36].

2.4 Метод спектрального віднімання

Реверберований сигнал це результат згортки чистого сигналу $s(t)$ та випадкової імпульсної характеристики приміщення $h(t)$. Також сигнал з реверберацією можна представити у вигляді доданків ранньої та пізньої складової.

$$y(t) = y_E(t) + y_L(t) = \int_{t=0}^{T_L} s(t-\tau)h(\tau)d\tau + \int_{T_L}^{T_A} s(t-\tau)h(\tau)d\tau, \quad (2.1)$$

де T_A – це довжина (в секундах) імпульсної характеристики приміщення $h(t)$,

T_L – це інтервал після якого починається пізня реверберація.

Для зменшення пізньої реверберації в спотвореному сигналі, короткотривалий спектр пізньої реверберації $Y_L(i, f)$ віднімається від короткотривалого спектру вхідного сигналу з реверберацією $Y(i, f)$. Короткотривалі спектри $Y_L(i, f)$ та $Y(i, f)$ це результат віконного перетворення Фур'є відповідних сигналів в часовому просторі $y_L(t)$ та $y(t)$, де i показує номер кадру, а f позначає частоту.

Очищений спектр сигналу $Y_{SS}(i, f)$, який відповідає вхідному сигналу без пізньої реверберації, може бути представлений як

$$Y_{SS}(i, f) = Y(i, f) - Y_L(i, f) = G_{SS}(i, f)Y(i, f). \quad (2.2)$$

Фільтр спектрального віднімання $G_{SS}(i, f)$, який визначений як

$$G_{SS}(i, f) = \frac{Y(i, f) - Y_L(i, f)}{Y(i, f)} = 1 - \frac{Y_L(i, f)}{Y(i, f)}. \quad (2.3)$$

Отже коротостроковий спектр пізньої реверберації має бути оцінений для того, щоби знайти оптимальне значення для фільтра спектрального віднімання [37].

2.5 Методи просторової обробки

Системи з одним мікрофоном використовують тільки спектральну і часову характеристику прийнятого сигналу. Реверберація і більшість зовнішніх джерел шуму, звичайно, мають просторове рознесення. Для того, щоби додатково використовувати цю особливість, мають бути використані декілька мікрофонів, і їх виходи мають бути об'єднані за допомогою відповідного просторового процесора. Процесор має складатися з ліній затримки і вагових коефіцієнтів для формування діаграми спрямованості [38] (рис. 2.4).

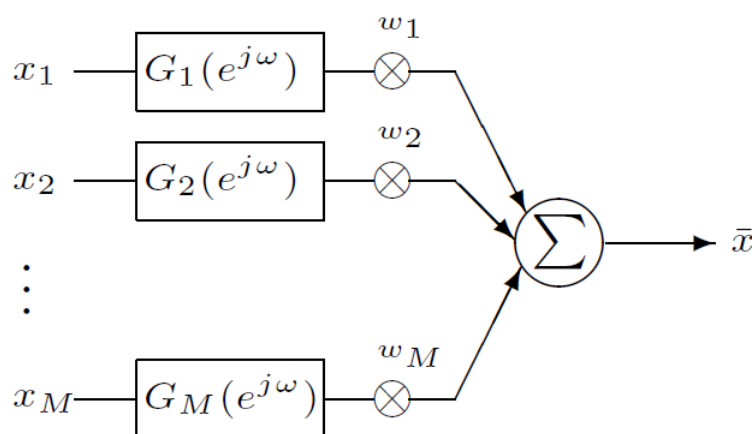


Рисунок 2.4 – Загальна структура формувача діаграми спрямованості

Методи з використанням мікрофонних решіток спрямовані на формування вузького променя діаграми спрямованості, де m -тий мікрофон

направляється в сторону джерела з використанням фазового управляючого фільтра $G_m(j\omega)$ і відповідного вагового коефіцієнта w_m .

Хоча просторові процесори дають значне поліпшення якості мови, придушення реверберації обмежено і придушення шуму недостатньо, коли шумове поле не є когерентним або дифузним [39].

2.6 Гомоморфна деконволюція

Цей метод базується на використанні кепстрального аналізу мовного сигналу. Комплексний кепстр має кілька властивостей, які роблять цей метод здатним для виконання деконволюції.

По-перше, сигнали, які в поєднанні згортою в часовій області, в комплексному кепстрі поєднанні адитивно. В результаті деконволюція зводиться до віднімання в кепстрів.

По-друге, комплексний кепстр є мірою варіації "сачтоти" (відомої як *quefrequency*) в логарифмічному спектрі, і тому сигнал, який повільно змінюється в логарифмічному спектрі, може бути відділений від швидко змінюючихся сигналів за допомогою віконної обробки комплексного кепстра (рис. 2.5).

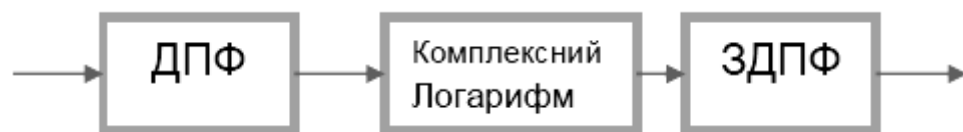


Рисунок 2.5 – Обчислення комплексного кепстру

Мова, як правило, повільно змінюється в логарифмічному спектрі і має комплексний кепстр, зосереджений близько нуля. Відлуння мають затримку відносно прямого сигналу, і можуть бути представлені за допомогою імпульсної характеристики, яка в логарифмічному спектрі характеризується швидкими "пульсаціями". Тому в комплексному кепстрі імпульсна характеристика складатиметься з імпульсів, зосереджених неподалік від нуля.

Розроблено процедури [40], в результаті яких комплексний кепстр обчислюється із сегментів ревербераційної мови, до якої застосовується експоненційно зважена віконна функція і компоненти косинусного перетворення, з відповідною імпульсною характеристикою будуть видалені.

Якщо комплексний кепстр ехо-сигналів має вигляд піків, вони визначаються відповідно до процедури пік-визначення і значення косинусного перетворення в точках їх виникнення встановлюються в нуль.

В той же час, обчислений кепстр множиться на віконну функцію, що призначена для збереження кепстра мови і видалення кепстра відлуння. Решта кепстра повторно перетворюється в часову область і множиться на зворотну експоненційну віконну функцію (рис. 2.6).

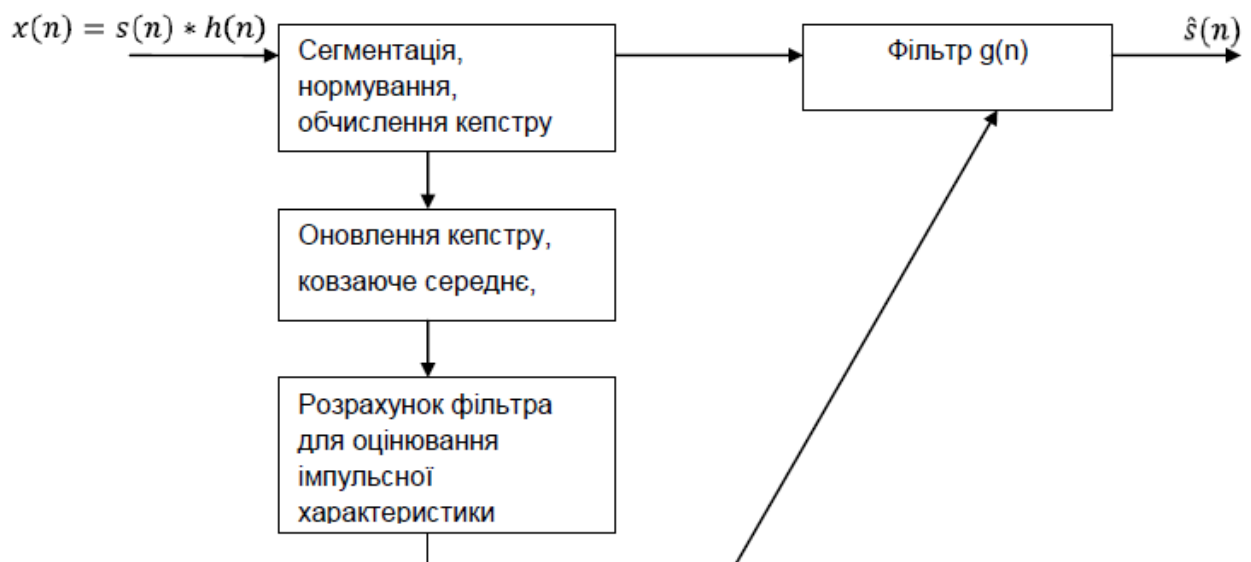


Рисунок 2.6 – Блок схема усунення реверберації на базі гомоморфної деконволюції

2.7 Висновок по розділу 2

На базі аналізу всіх алгоритмів усунення реверберації можна зробити висновок, що найбільш придатним для подальшого використання є алгоритм спектрального віднімання. Оскільки в нашому випадку ми не маємо достатньої інформації про джерело сигналу і тільки часткові знання про канал поширення сигналу.

3 МЕТОД СПЕКТРАЛЬНОГО ВІДНІМАННЯ І ЙОГО РІЗНОВИДИ

3.1 Оцінка спектру реверберованого сигналу

Метод спектрального віднімання базується на тому що реверберований сигнал представляється за допомогою статистичного моделювання реверберації в приміщенні з використанням моделі Полака [41]. Дана модель описана в розділі 2.

Реверберований сигнал це результат згортки чистого сигналу $s(t)$ та випадкової імпульсної характеристики приміщення $h(t)$

$$x(t) = \int_{-\infty}^{\infty} s(t)h(t - \tau)d\tau. \quad (3.1)$$

Згідно моделі Полака

$$h(t) = b(t)e^{-\alpha t}, \text{ для } t \geq 0. \quad (3.2)$$

де $b(t)$ – білий гаусівський стаціонарний шум з нульовим середнім.

Тоді згортка може бути представлена у вигляді

$$x(t) = e^{-\alpha t} \int_{-\infty}^{\infty} s(t)b(t - \tau)e^{\alpha\tau} d\tau. \quad (3.3)$$

тому що при $t - \tau \geq 0$, $\tau \leq t$, що дає верхню границю інтеграла.

Автокореляційна функція $r_{xx}(t, t + \tau) = E\{x(t)x(t, t + \tau)\}$ реверберованого сигналу x в момент часу та зі зсувом τ

$$r_{xx}(t, t + \tau) = E \left\{ \int_{-\infty}^t \int_{-\infty}^{t+\tau} s(\theta) s(\theta') h(t - \theta) h(t + \tau - \theta') d\theta d\theta' \right\}, \quad (3.4)$$

де $E\{\}$ позначає математичне сподівання (expectation).

З використанням статистичної моделі Полака автокореляційну функцію можна записати у вигляді

$$r_{xx}(t, t + \tau) = E \left\{ \int_{-\infty}^t \int_{-\infty}^{t+\tau} s(\theta) s(\theta') b(t - \theta) b(t + \tau - \theta') e^{-\alpha t} e^{-\alpha(t+\tau)} e^{\alpha\theta} e^{\alpha\theta'} d\theta d\theta' \right\}. \quad (3.5)$$

Тобто, лінійна операція згортки застосовується до сигналів $s(t)$ і $b(t)$.

Очікуване значення інтеграла дорівнює інтегралу від очікуваного значення, отже

$$r_{xx}(t, t + \tau) = \int_{-\infty}^t \int_{-\infty}^{t+\tau} E\{s(\theta) s(\theta') b(t - \theta) b(t + \tau - \theta')\} e^{-\alpha t} e^{-\alpha(t+\tau)} e^{\alpha\theta} e^{\alpha\theta'} d\theta d\theta'. \quad (3.6)$$

Якщо вважати s і b незалежними випадковими процесами, то можна записати

$$\begin{aligned} r_{xx}(t, t + \tau) &= \int_{-\infty}^t \int_{-\infty}^{t+\tau} E\{s(\theta) s(\theta')\} E\{b(t - \theta) b(t + \tau - \theta')\} e^{-\alpha t} e^{-\alpha(t+\tau)} e^{\alpha\theta} e^{\alpha\theta'} d\theta d\theta' = \\ &= e^{-2\alpha t} \int_{-\infty}^t \int_{-\infty}^{t+\tau} E\{s(\theta) s(\theta')\} E\{b(t - \theta) b(t + \tau - \theta')\} e^{\alpha(\theta + \theta' - \tau)} d\theta d\theta'. \end{aligned} \quad (3.7)$$

Використовуючи статистичну модель імпульсної характеристики приміщення, описану виразом (3.2)

$$E\{s(\theta) s(\theta')\} E\{b(t - \theta) b(t + \tau - \theta')\} = \sigma^2 \delta(\theta - \theta' + \tau), \quad (3.8)$$

де δ – позначає функцію Дірака

σ^2 – дисперсія процесу $b(t)$.

Тоді

$$\begin{aligned}
 r_{xx}(t, t + \tau) &= e^{-2\alpha t} \int_{-\infty}^t E\{s(\theta)s(\theta + \tau)\} e^{\alpha(\theta + \theta' - \tau)} \sigma^2 e^{2\alpha\theta} d\theta = \\
 &= e^{-2\alpha t} \int_{t-T}^t E\{s(\theta)s(\theta + \tau)\} \sigma^2 e^{2\alpha\theta} d\theta + \\
 &+ e^{-2\alpha t} \int_{-\infty}^{t-T} E\{s(\theta)s(\theta + \tau)\} \sigma^2 e^{2\alpha\theta} d\theta. \quad (3.9)
 \end{aligned}$$

Автокореляційна функція в момент t може бути розділена на дві частини. Перша частина залежить від прямого сигналу в часовому інтервалі $t-T$ і t , в той час як друга частина залежить від сигналу реверберації і відповідає за пізні відбиття. Розглянемо просторово усереднену автокореляцію в момент $t-T$

$$r_{xx}(t-T, t-T + \tau) = e^{-2\alpha(t-T)} \int_T^t E\{s(\theta)s(\theta + \tau)\} \sigma^2 e^{2\alpha\theta} d\theta. \quad (3.10)$$

Отже, автокореляція в момент часу t може бути виражена як

$$r_{xx}(t, t + \tau) = r_{x_d x_d}(t, t + \tau) + r_{x_r x_r}(t, t + \tau), \quad (3.11)$$

де

$$\begin{aligned}
 r_{x_d x_d}(t, t + \tau) &= e^{-2\alpha t} \int_T^t E\{s(\theta)s(\theta + \tau)\} \sigma^2 e^{2\alpha\theta} d\theta, \\
 r_{x_r x_r}(t, t + \tau) &= e^{-2\alpha T} r_{xx}(t-T, t-T + \tau). \quad (3.12)
 \end{aligned}$$

На практиці сигнали можна розглядати як стаціонарні протягом періоду часу, який малий в порівнянні з часом реверберації T_r . Це виправдано тим, що експонентний спад дуже повільний, і що мова є квазістаціонарною.

Нехай T_s є проміжок часу протягом якого сигнал мови можна вважати стаціонарним, який, близький до 20...40 мс [42]. Будемо вважати, що $T_s < T < T_r$.

Спектральна характеристика випадкового сигналу отримується шляхом обчислення перетворення Фур'є від автокореляційної функції. Тобто, розподіл потужності за частотою задається виразом:

$$Y(f) = \int_{-\infty}^{\infty} r(\tau) e^{-2\pi f\tau} d\tau. \quad (3.13)$$

Короткочасна спектральна густина приблизно рівна:

$$Y_{xx}(t, f) = Y_{x_d x_d}(t, f) + Y_{x_r x_r}(t, f),$$

$$Y_{x_r x_r}(t, f) = e^{-2\alpha T} Y_{xx}(t, f). \quad (3.14)$$

Таким чином, ми можемо оцінити спектральну щільність потужності прямого сигналу за допомогою спектрального віднімання спектральної щільності потужності пізньої реверберації.

3.2 Метод Лебарта, Бочера і Дебніха

Для того щоб отримати амплітуду короткострокового спектру пізньої реверберації, можна обчислити спектральну щільність потужності (Power Spectral Density – PSD) пізньої реверберації та взяти квадратний корінь з неї [43]. Використовуючи той факт, що PSD – це віконне перетворення Фур'є

(STFT) від автокореляційної функції пізньої реверберації $y_L(t)$ в часовій області $E[y_L(t)y_L(t+\tau)]$. Суттєвим є той факт, що для того, щоб оцінити PSD або автокореляцію складової пізньої реверберації, більш детальна інформація про більш пізню та ранню частини реверберації вхідного сигналу $y(t)$ не потрібна.

Полак ввів статистичну модель [44] імпульсної характеристики приміщення (Acoustic Impulse Response – AIR), яка заснована на послідовності незалежних і однаково розподілених випадкових величин $b(t)$ з нульовим математичним очікуванням та нормальним законом розподілу:

$$h(t) = b(t)e^{-\rho t}, \text{ при } t \geq 0. \quad (3.15)$$

де ρ – це постійна середнього затухання, яка пов'язана з часом реверберації T_r виразом

$$\rho = \frac{3 \ln(10)}{T_r}. \quad (3.16)$$

Зі стаціонарним гаусівським шумом з нульовим середнім і дисперсією σ^2 – $b(t) \sim N(0, \sigma^2)$, який базується на імпульсному відгуку AIR, автокореляційна функція вхідного сигналу може бути записана як

$$\begin{aligned} E\{y(t)y(t+\tau)\} &= \int_{t=0}^t \int_{t=0}^{t+\tau} E\{s(\theta)s(\theta')\} E\{b(t-\theta)b(t+\tau-\theta')\} \cdot e^{\rho(\theta+\theta'-\tau)} d\theta d\theta' = \\ &= e^{-2\rho t} \int_{t=0}^t E\{s(\theta)s(\theta+\tau)\} \cdot \sigma^2 e^{2\rho\theta} d\theta. \end{aligned} \quad (3.17)$$

Взявши припущення що енергія відбиттів більша ніж енергія прямого звуку, автокореляція вхідного сигналу може бути пов'язана з автокореляцією пізніх відбиттів. За допомогою розкладання $y(t)$ на $y_E(t)$ та $y_L(t)$ через

визначений часовий інтервал T_L , можна показати що пізня реверберація може бути змодельована некорельованим шумовим процесом. Як наслідок автокореляційна функція пізньої реверберації сигналу виглядає наступним чином

$$E\{y_L(t)y_L(t+\tau)\} = e^{-2\rho T_L} E\{y(t-T_L)y(t-T_L+\tau)\}. \quad (3.18)$$

З цих міркувань можна, нарешті, вивести оцінку для PSD пізньої реверберації $|Y_L(j, f)|^2$, що відповідає короткостроковому представленню виведеної автокореляції:

$$|Y_L(j, f)|^2 = e^{-2\rho T_L} |\hat{Y}(j - N_L, f)|^2. \quad (3.14)$$

Оскільки це короткостроковий PSD, він більше не є безперервним у часі, але він організований в кадри (відрізки), тому N_L замінює T_L . Оцінка PSD реверберації вхідного сигналу залежить не тільки від поточного кадру $|Y_L(j, f)|^2$, але також від оцінки PSD попереднього кадру. Вона може бути отримана з виразу для PSD через автокореляційну функцію. Оцінка PSD вхідного сигналу може бути отримана рекурсивно через середнє між поточним та попереднім кадром. Також застосовується згладжування з коефіцієнтом $0 \leq \alpha \leq 1$:

$$|\hat{Y}(j, f)|^2 = \alpha \cdot |\hat{Y}(j - N_L, f)|^2 + (1 - \alpha) \cdot |Y(j, f)|^2. \quad (3.15)$$

Як було зазначено вище, повний короткостроковий спектр пізньої реверберації $Y_L(j, f)$ можна скласти з його амплітуди і фази. Оскільки фаза спектру пізньої реверберації така сама як і вхідного реверберованого

сигналу, спектральне віднімання не змінює фазу. Тому, АЧХ фільтра спектрального віднімання може бути представлена як

$$G_{SS}(j, f) = 1 - \frac{|Y_L(j, f)|}{|Y(j, f)|} = 1 - \frac{e^{-2\rho T_L} |\hat{Y}(j - N_L, f)|}{|Y(j, f)|}. \quad (3.16)$$

На рис.3.1. показана структурна схема фільтра спектрального віднімання Лембарта.

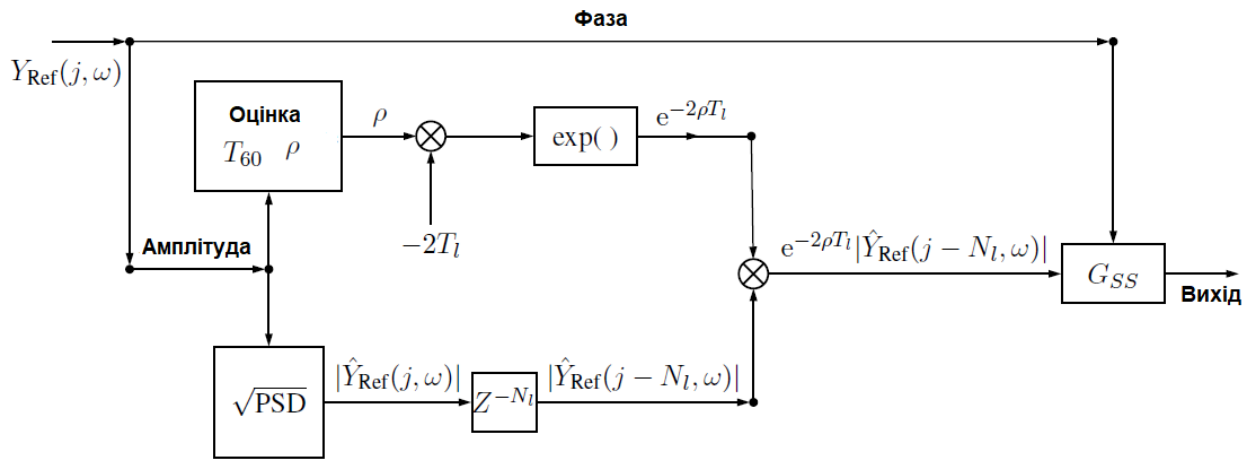


Рисунок 3.1 – Схема фільтра спектрального віднімання Лембарта

3.3 Метод Ву та Ванга

Метод Ву та Ванга описує короткостроковий спектр потужності частини сигналу з пізньою реверберацією як згладжений і зсунутий спектр PSD сигналу з реверберацією. Згладжування в часовій області відповідає за ефект розмиття пізніх імпульсів в ревербераційному відгуку. Більше того, PSD пізніх компонентів розглядається як згладжена і зсунута версія спектру PSD реверберованої мови.

$$|Y_L(j, f)|^2 = \gamma r(j - \rho) \cdot |\hat{Y}(j, f)|^2. \quad (3.15)$$

Зсув представлений параметром затримки ρ та коефіцієнтом масштаба γ який визначає відносну силу пізньої реверберації у короткостроковому

спектрі PSD. Зсув і згладжування можна описати згорткою спектру PSD сигналу реверберації та асиметричної функції згладжування, яка нагадує форму вирівняної ревербераційної імпульсної характеристики. Даним умовам відповідає релєєвський розподіл, який визначається як

$$r(j) = \begin{cases} \frac{j+a}{a^2} e^{-\frac{(j+a)^2}{a^2}}, & f \geq -a \\ 0, & \text{інакше} \end{cases} \quad (3.17)$$

На рис. 3.2 показано графік функції Релєя $\gamma r(j-\rho)$ для параметрів $\rho = 300$, $\alpha = 200$ і $\gamma = 3000$.

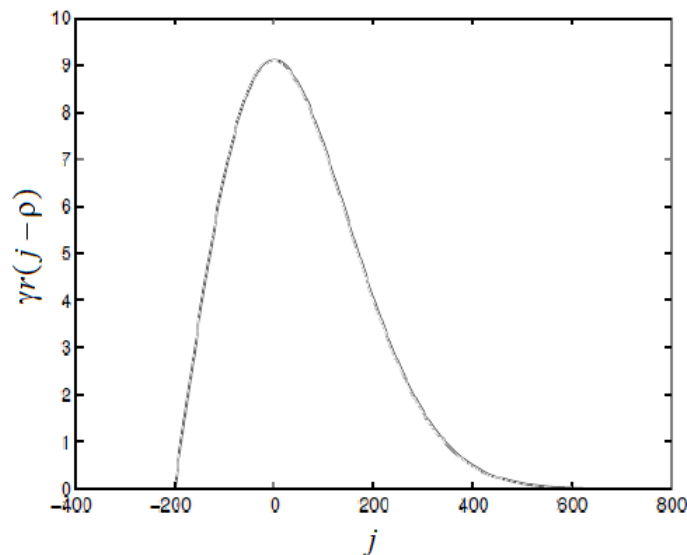


Рисунок 3.2 – Графік функції згладжування $\gamma r(j-\rho)$ для параметрів $\rho = 300$, $\alpha = 200$ і $\gamma = 3000$

Безревербераційний спектр сигналу $Y_{SS}(j, f)$ може бути визначений шляхом застосування коригуючого еквалайзера $G_{SS}(j, f)$, характеристика якого рахується методом Ву і Ванга зі спектру вхідного сигналу з реверберацією:

$$Y_{SS}(j, f) = Y(j, f) - Y_L(j, f) =$$

$$= \left\{ \frac{|\hat{Y}(j, f)|^2 - |Y_L(j, f)|^2}{|\hat{Y}(j, f)|^2} \right\} Y(j, f) = G_{SS}(j, f) Y(j, f). \quad (3.18)$$

На рис.3.3 показана структурна схема фільтру спектрального віднімання за методом Ву і Ванга.

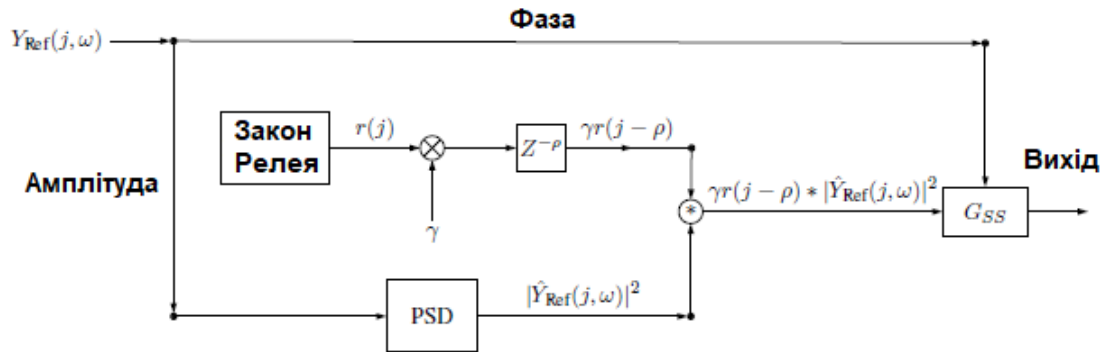


Рисунок 3.3 – Структурна схема фільтру спектрального віднімання за методом Ву і Ванга

3.4 Метод Блейкера та Георганті

До отримання потрібних результатів може призвести не тільки згортка функції релеєвського розподілу зі спектром сигналу з реверберацією. Подібний результат може дати також згортка цієї функції згладжування з реверберованим сигналом в часовій області, що використовується для імітації ефекту розмиття пізньою реверберацією. Згладжування і зсув в часовій області для реверберації тепер визначається згладженою версією реверберованого сигналу з декількох попередніх кадрів.

$$y_L(j) = \gamma r(j - \rho) * y(j), \quad (3.19)$$

де * позначає операцію згортки.

Це згладжування і зсув часового представлення реверберації мовлення можуть бути представлені в частотній області

$$|Y_L(j, f)|^2 = \gamma \cdot STFT\{\gamma r(j - \rho)\} \cdot |\hat{Y}(j, f)|^2. \quad (3.20)$$

Згортка в часовій області відповідає множенню в частотній області. Множення спектру ревербованого сигналу виконується з віконним перетворенням Фур'є із функцією затриманого релеєвського розподілу $STFT\{\gamma r(j - \rho)\} = R(f)$.

На рис.3.4 показана структурна схема фільтру спектрального віднімання за методом Ву і Ванга з модифікацією Блейкера та Георганті.

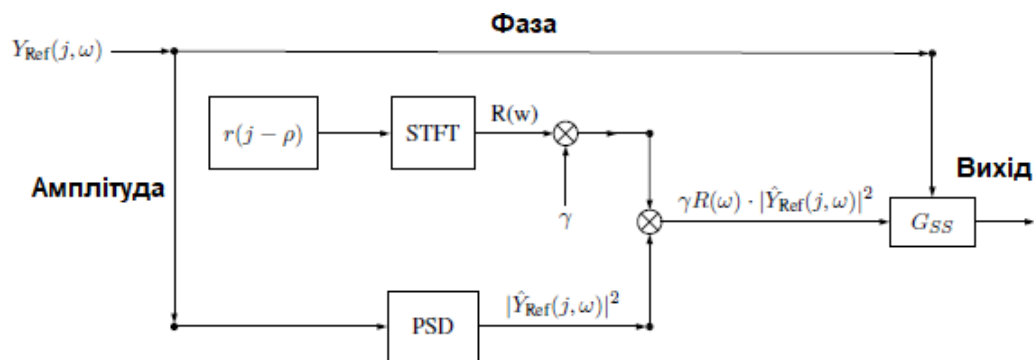


Рисунок 3.4 – Структурна схема фільтру спектрального віднімання за методом Ву і Ванга з модифікацією Блейкера та Георганті

3.5 Метод Фарая та Катаоки

В методі Фарая та Катаоки розраховується функція корекції частотної характеристики фільтру спектрального віднімання безпосередньо за допомогою оцінки PSD складової пізньої реверберації та вхідного сигналу із реверберацією.

$$G_{ss}(j, f) = \frac{|\hat{Y}(j, f)|^2 - |Y_L(j, f)|^2}{|\hat{Y}(j, f)|^2}. \quad (3.20)$$

В методі пропонується оцінювати короточасний спектр PSD пізньої реверберації в кожному кадрі шляхом підсумовування відфільтрованих,

старих версій кадрів реверберованого сигналу [45]. Спочатку розраховуються вагові коефіцієнти кожного кадру спектру PSD, а потім спектр потужності попередніх кадрів, які додаються з відповідними ваговими коефіцієнтами.

$$|Y_L(j, f)|^2 = \sum_{l=1}^K |a_l(j, f)|^2 \cdot |\hat{Y}(j-l, f)|^2. \quad (3.21)$$

Вагові коефіцієнти роблять можливим вплив попереднього кадру спектру PSD сигналу з реверберацією на спектр поточного кадру. Чим більшу схожість показують два PSD, тим сильнішим є вплив їх попереднього спектру на поточний кадр для складової пізньої реверберації.

$$a_l(j, f) = E \left\{ \frac{\hat{Y}(j, f) \cdot \hat{Y}^*(j-l, f)}{|\hat{Y}(j-l, f)|^2} \right\}. \quad (3.20)$$

На рис.3.5 показана структурна схема фільтру спектрального віднімання за методом Фарая та Катаоки.

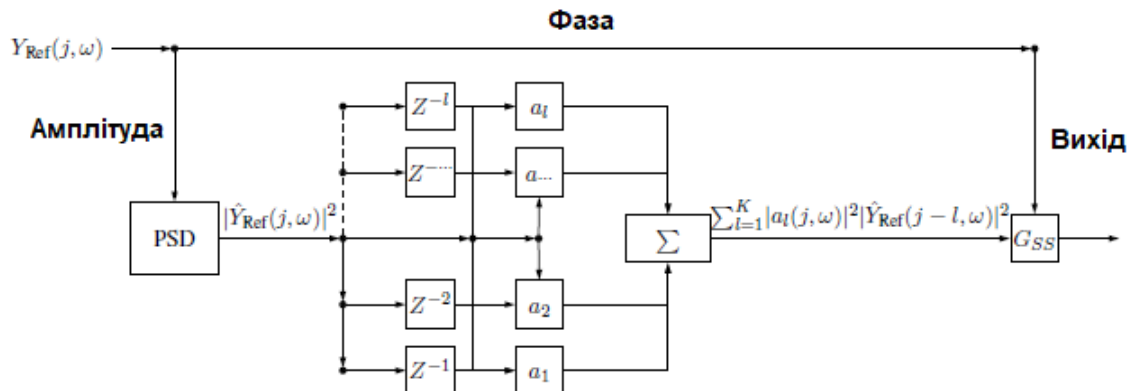


Рисунок 3.5 – Структурна схема фільтру спектрального віднімання за методом Фарая та Катаоки

3.6 Висновок по розділу 3

Було розглянуто підходи, що дозволяють оцінити спектр пізньої реверберації, для алгоритму спектрального віднімання.

Звичайно, кожен з підходів має свої переваги та недоліки, але для нашої роботи більш підходить метод Лебарта, оскільки він використовує час реверберації для розрахунків, що дозволить підняти точність вимірювання при відомих параметрах приміщення. Також це дозволить більш детально проаналізувати якість роботи алгоритму, оскільки ми зможемо за допомогою генераторів штучної реверберації задавати час реверберації та досліджувати результати фільтрації. Тому в подальшому будемо використовувати метод спектрального віднімання з оцінкою спектра пізньої реверберації за методом Лебарта.

4 ДОСЛІДЖЕННЯ ДЕРЕВЕРБЕРАЦІЇ ЗВУКОВОГО СИГНАЛУ МЕТОДОМ СПЕКТРАЛЬНОГО ВІДНІМАННЯ

4.1 Умови експерименту

Моделювання виконувалося в програмному середовищі MATLAB. Для моделювання застосовано алгоритм подавлення пізньої реверберації на базі спектрального віднімання з використанням методу Лебарта для оцінки спектру пізньої реверберації.

Для отримання реверберованого сигналу використовувався алгоритм Мура на основі мережі ліній затримок.

Моделювання виконувалося за таким алгоритмом (рис 4.1):

- вхідний сигнал подається на генератор штучної реверберації на виході отримуємо сигнал з реверберацією;
- сигнал з реверберацією подається на схему усунення пізньої реверберації, після чого отримуємо відфільтрований сигнал;
- порівнюємо вхідний чистий сигнал та відфільтрований сигнал після роботи алгоритму за допомогою спектрограм та середньоквадратичної різниці.

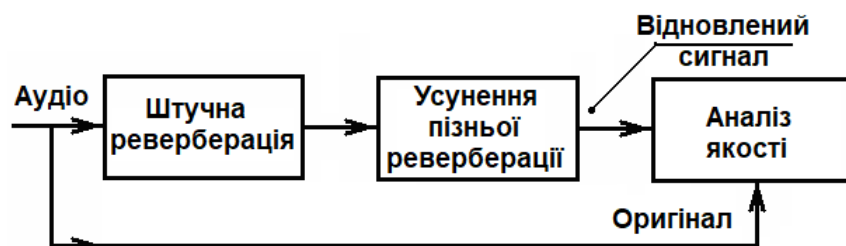


Рисунок 4.1 – Схема модельного експерименту

Також слід зазначити, що дуже важко оцінити якість роботи алгоритму, оскільки ми фільтруємо тільки пізню реверберацію. А наявність ранньої реверберації вже не дасть змоги в повній мірі порівнювати вихідний відфільтрований сигнал з чистим вхідним сигналом.

Лістинг алгоритму програми для усунення реверберації наведено в Додатку А. Лістинг програми для генерації штучної реверберації – у Додатку Б.

4.2 Моделювання гармонічного сигналу

Спочатку, для більшої наглядності роботи алгоритму, проведемо моделювання на простому сигналі з декількома гармоніками (гітарному акорді), з великим часом реверберації ($T_r = 1,5$ с).

Як видно на рис. 4.2 реверберований сигнал має „хвіст“ пізньої реверберації, а відновлений сигнал має вже подавлену пізню реверберацію. Тому можна зробити висновок, що реалізація алгоритму є працездатною.

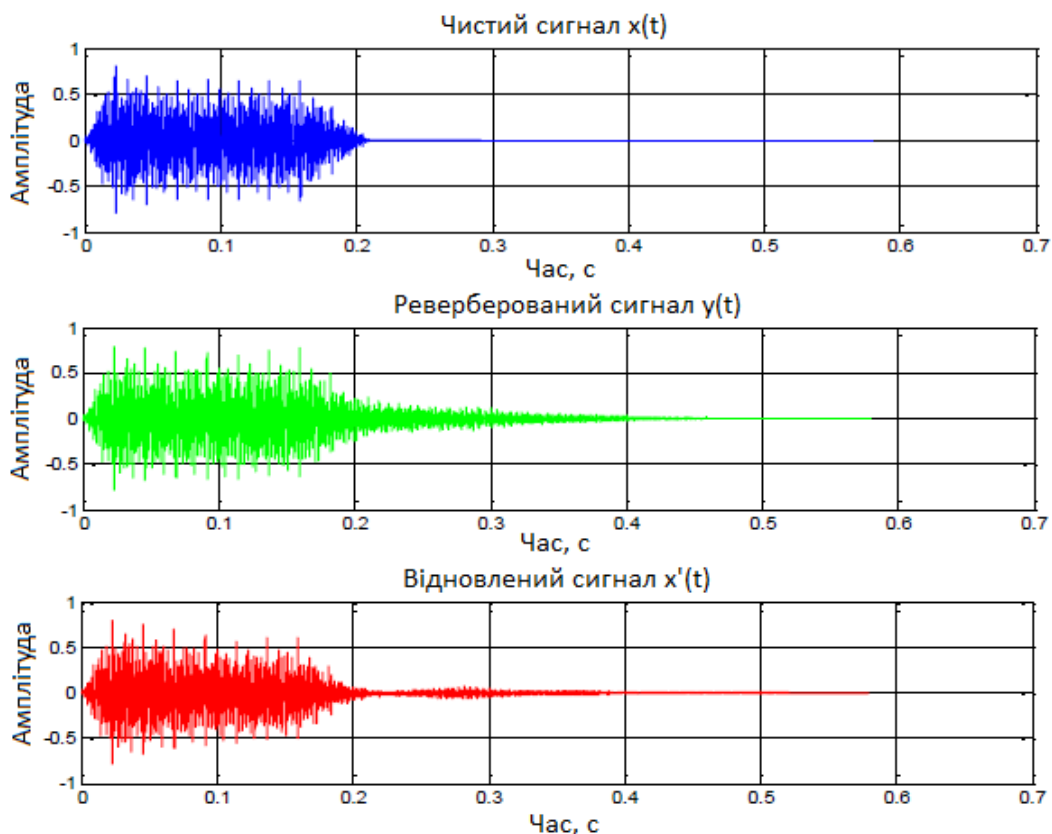


Рисунок 4.2 – Чистий, реверберований і відновлений сигнали

Аналогічну картину можна спостерігати на рис. 4.3, де показані спектрограми чистого, реверберованого і відновленого сигналів.

На спектрограмі видно, що протяжність пізньої реверберації на різних частотах є різною. Тобто, імпульсна характеристика приміщення має чітко виражену частотну залежність. В акустиці це явище називається частотною залежністю часу реверберації. Це явище обумовлене різним коефіцієнтом поглинання покриттів приміщення $a(f)$ на різних частотах.

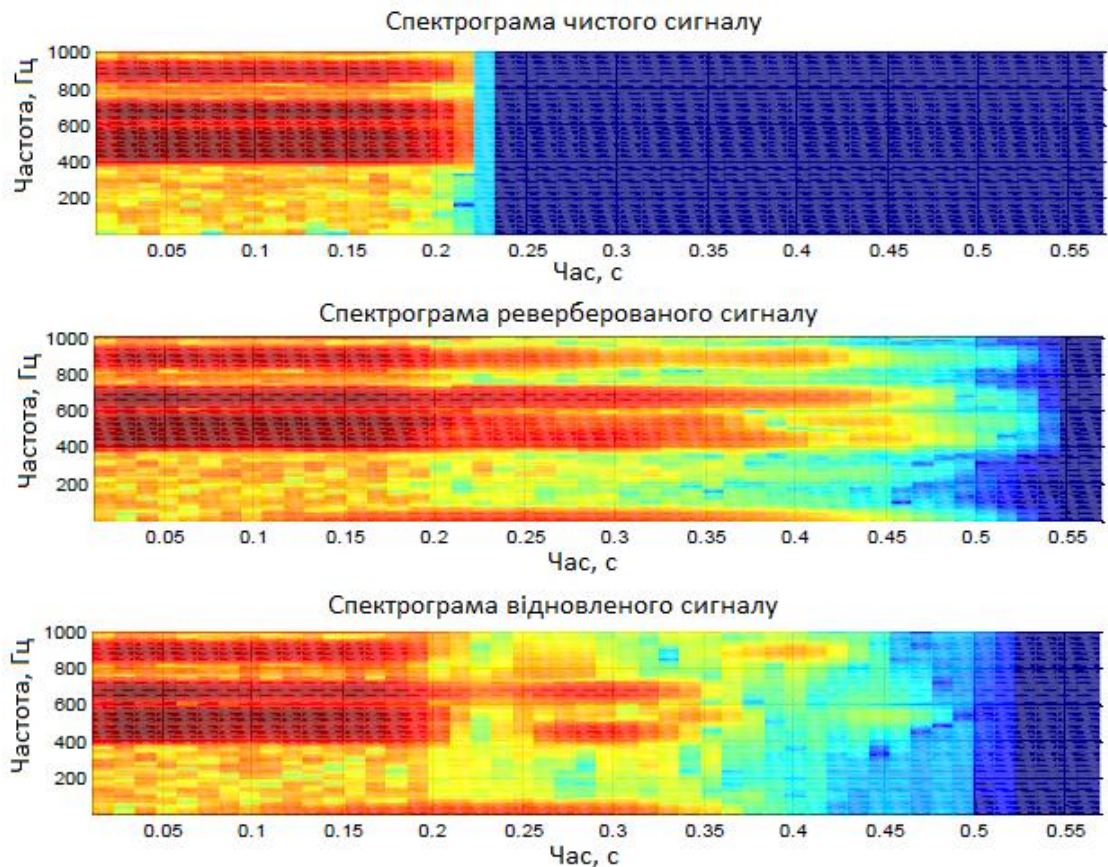


Рисунок 4.3 – Спектрограма чистого, реверберованого і відновленого сигналів

На рис.4.3 для прикладу показана частотна залежність коефіцієнта звукопоглинання для поролонового звукопоглинача (червона лінія – поролонова пірамідка 75 мм, синя лінія – поролонова хвиля 50 мм).

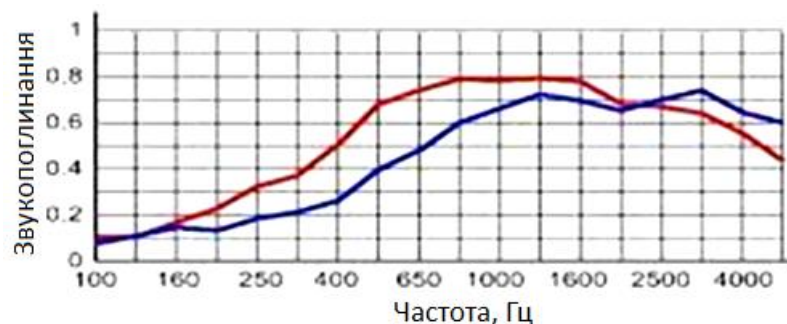


Рисунок 4.3 – Частотна залежність коефіцієнта звукопоглинання (червоне – поролонова пірамідка 75 мм, синє – поролонова хвиля 50 мм)

Отже, можна зробити висновок, що реалізація алгоритму є працездатною.

4.3 Моделювання з використанням мовного сигналу

Розглянемо роботу алгоритму для випадку реального мовного сигналу. Для створення штучної реверберації будемо використовувати згортку чистого сигналу з реальною імпульсною характеристикою приміщення. Файл імпульсного відгуку приміщення завантажений з Інтернет-ресурсу.

Результати даного експерименту показані на рис. 4.4.

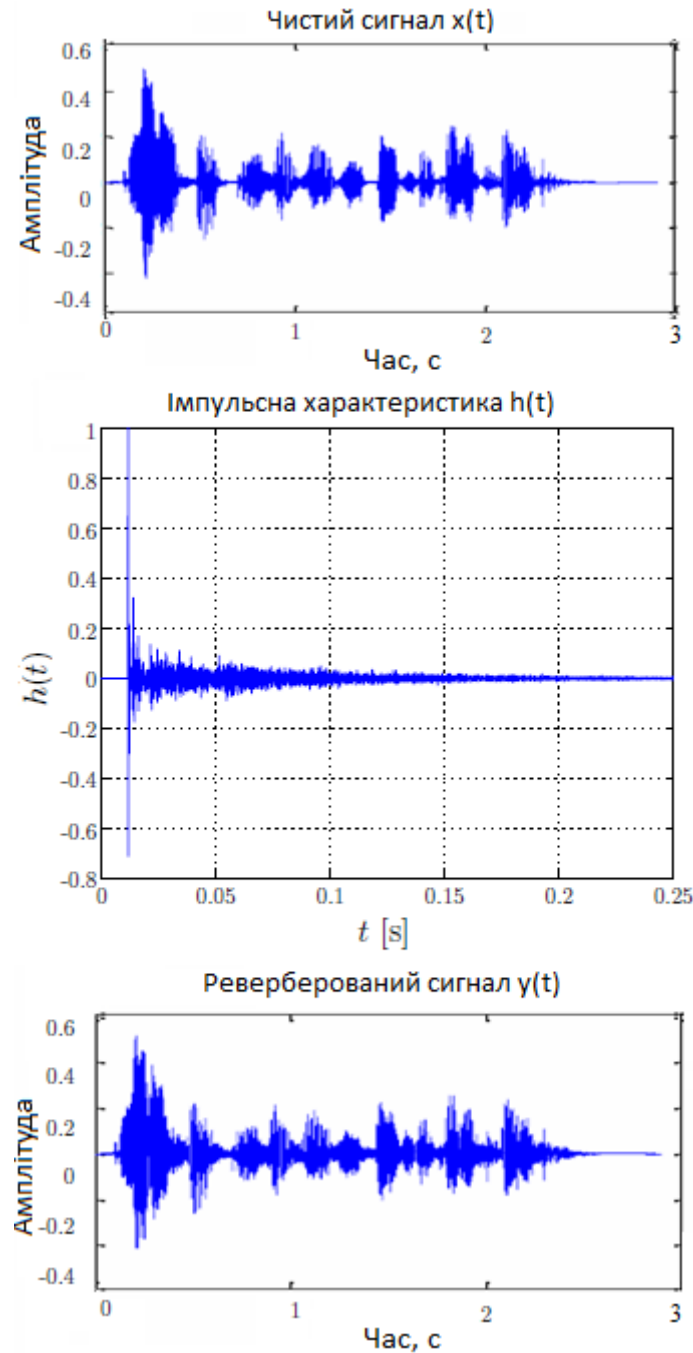


Рисунок 4.4 – Моделювання з використанням мовного сигналу

Відфільтрований сигнал показаний на рис. 4.5, а спектрограми чистого, реверберованого та відновленого сигналів – на рис.4.6.

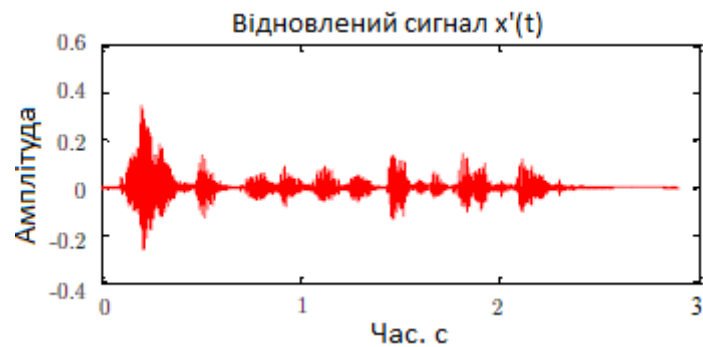


Рисунок 4.5 – Відновлений мовний сигнал

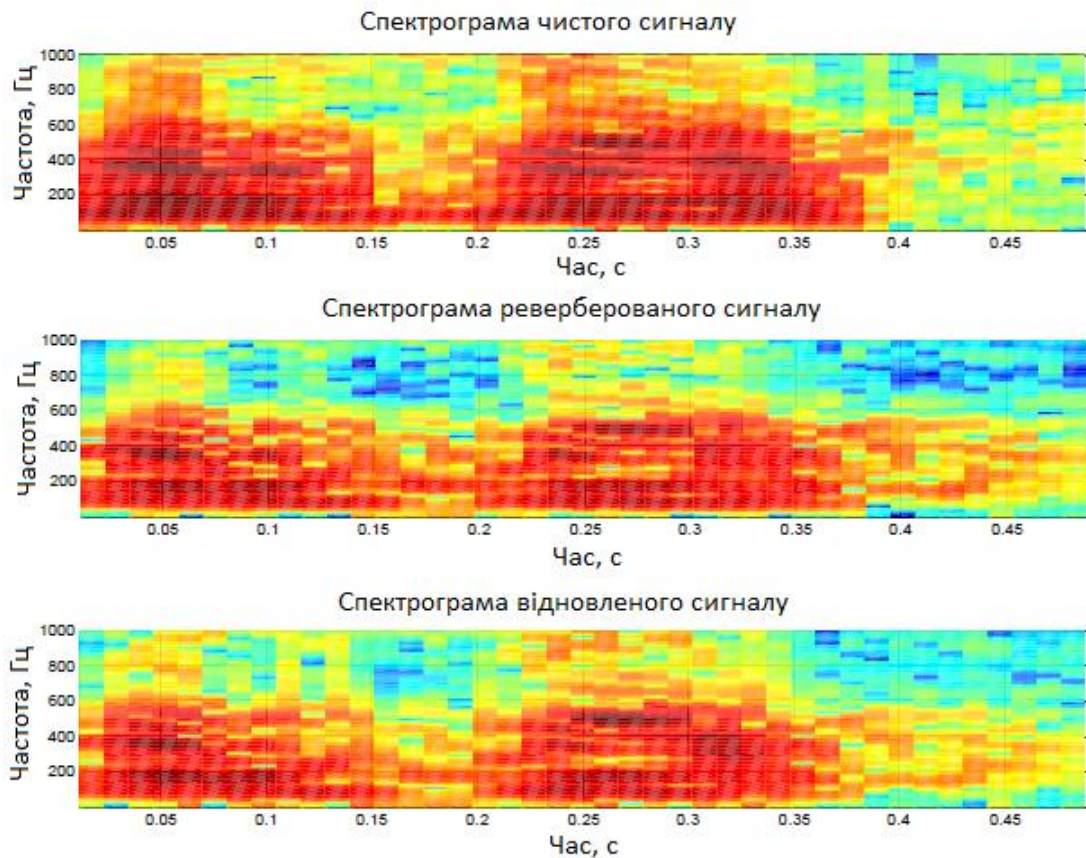


Рисунок 4.6 – Спектрограма чистого, реверберованого і відновленого мовного сигналів

Проведемо аналіз роботи алгоритму в залежності від зміни вхідних параметрів.

4.4 Аналіз роботи алгоритму дереверберації в залежності від вхідних параметрів

Було проведено аналіз роботи алгоритму дереверберації в залежності від зміни вхідних параметрів. Дослідження проводились з використанням мовного сигналу та генератора штучної реверберації Мура. Для оцінки різниці між чистим сигналом без реверберації $x(t)$ та відфільтрованим сигналом після алгоритму дереверберації $\hat{x}(t)$ використовувалося усереднене по тривалості сигналу T середньоквадратичне відхилення

$$RMS = \sqrt{\sum_{t=1}^T \frac{(x(t) - \hat{x}(t))^2}{T}}. \quad (4.1)$$

Слід зазначити, що завжди буде якийсь мінімальний рівень RMS оскільки в сигналі все ще присутня рання реверберація.

На рис.4.7 показана залежність RMS від порядку моделі штучного ревербератора.

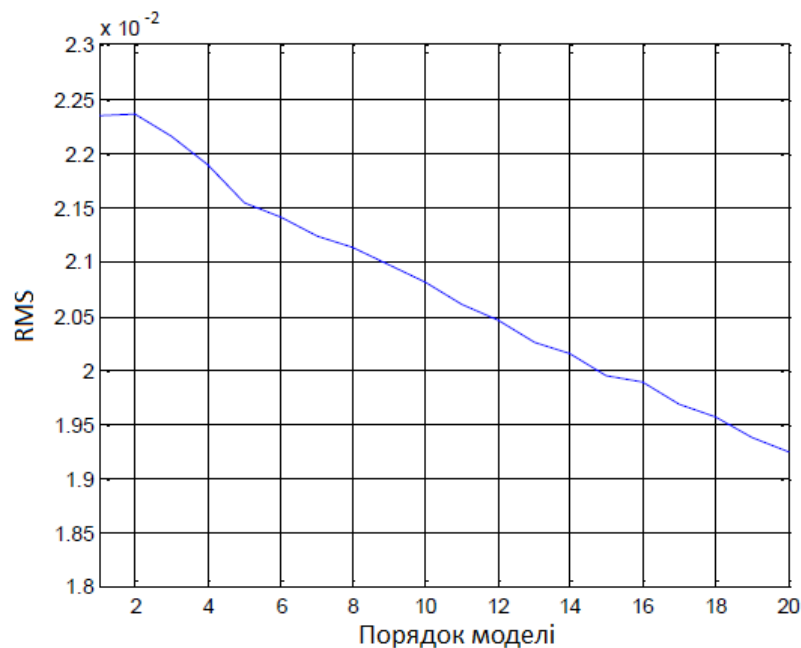
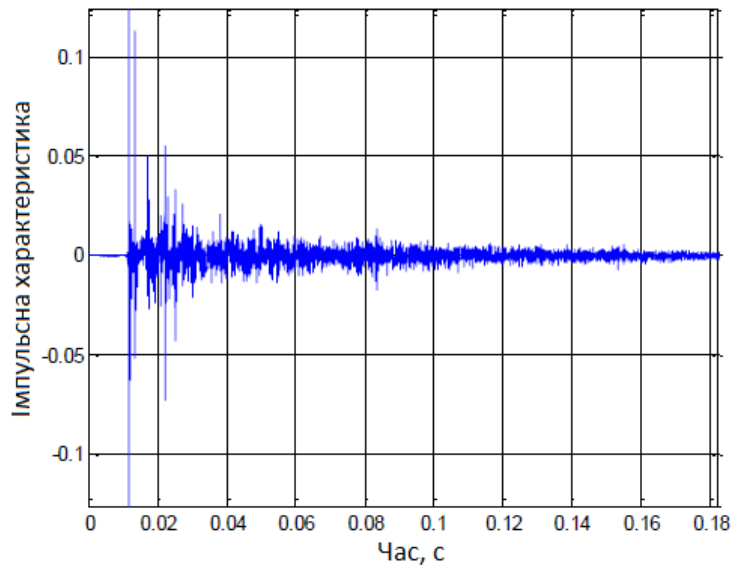
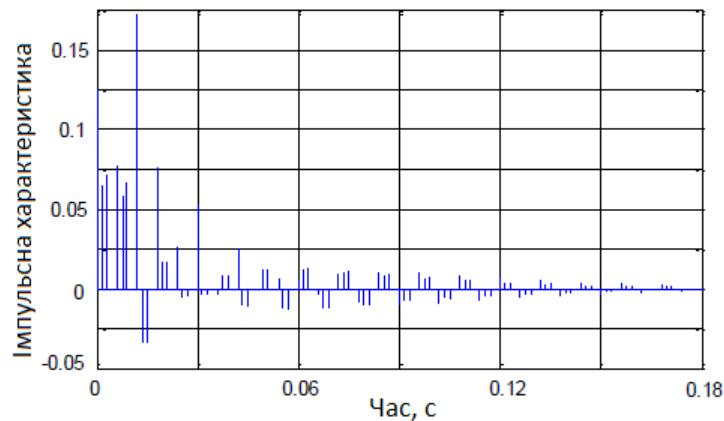


Рисунок 4.7 – Залежність RMS від порядку моделі штучного ревербератора

Щоби пояснити дану криву, розглянемо імпульсні відгуки ревербератора при різних порядках моделі. Рис.4.8, а відповідає імпульсній характеристиці при порядку моделі 20, рис. 4.8, б – при порядку моделі 5.



а)



б)

Рисунок 4.8 – Імпульсні відгуки ревербератора при різних порядках моделі: а – 20, б – 5

Отже, як видно з результатів моделювання, імпульсні характеристики мають однаковий характер, але вони по різному створюють ранні та пізні ревербераційні ефекти. В моделі Мура з порядком 5 імпульсна реакція звучить надто дискретно, що призводить до зернистості звуку, а в моделі з порядком 20 пізні відбиття більш розсіяні.

Отже, метод дереверберації методом спектрального віднімання гірше працює з імпульсними звуковими сигналами і краще – з сигналами, що

мають плавну атаку. Тобто, саме з мовними сигналами, метод спектрального віднімання працює краще.

На рис. 4.9 показана залежність RMS від коефіцієнта звукопоглинання поверхонь в моделі штучного ревербератора.

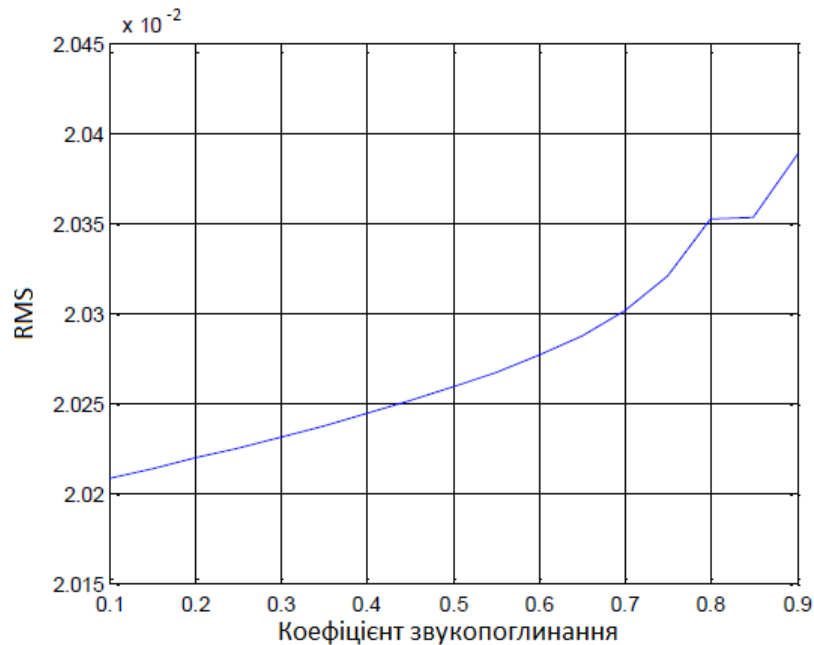


Рисунок 4.9 – Залежність RMS від коефіцієнта звукопоглинання поверхонь

Як слідує з рис.4.9, при збільшенні звукопоглинання RMS хоч і не значно, але зростає. На перший погляд, це парадоксальний результат. Але давайте проаналізуємо цей ефект детально.

При зростанні коефіцієнта звукопоглинання кількість пізніх відбиттів різко зменшується, а кількість ранніх відбиттів залишається приблизно на тому ж рівні. Алгоритм спектрального віднімання в цьому випадку призначений боротися саме з пізніми відбиттями. При їх невеликій кількості алгоритм більш спотворює сигнал, ніж подавляє відбиття. Пізня реверберація дозволяє алгоритму набагато точніше оцінити імпульсну характеристику приміщення.

Отже алгоритм спектрального віднімання доцільно застосовувати тільки при сильній реверберації і високому рівні пізніх відбиттів.

На рис. 4.10 показана залежність RMS від об'єму приміщення в моделі штучного ревербератора.

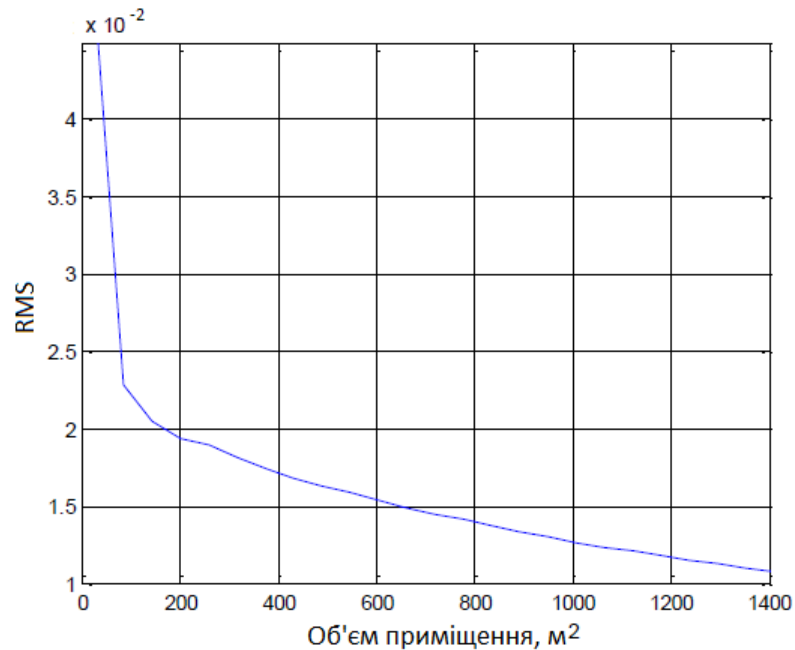


Рисунок 4.10 – Залежність RMS від об'єму приміщення

Дана залежність підтверджує висновок, що алгоритм спектрального віднімання ефективніше працює при тривалій реверберації і високому рівні пізніх відбиттів. Тому що при збільшенні об'єму приміщення пізні відбиття стають більш вираженими, що дозволяє алгоритму точніше визначити імпульсну характеристику приміщення.

4.5 Висновок по розділу 4

Було проведено моделювання роботи алгоритму, як на штучних, так і на реальних сигналах. На результатах видно, що алгоритм усуває пізню реверберацію. Складно об'єктивно оцінити якість дереверберації, оскільки в сигналі залишається рання реверберація.

Проведено аналіз роботи алгоритму в залежності від зміни вхідних параметрів важливих для реверберації – порядок моделі ревербератора, коефіцієнт звукопоглинання поверхонь, об'єм приміщення – що дало змогу оцінити роботу алгоритму в різних умовах.

ВИСНОВКИ

Процес звукозапису найчастіше будується за схемою, щоби реверберації не виникало. Професійні записи голосу відбуваються у спеціальній ізолюваній від шуму і відбиттів акустичній кабіні з малою реверберацією.

Дані умови виконуються не завжди вдається виконати, і у руки звукорежисера часто потрапляють фонограми, зроблені в непристосованих умовах. Це можуть бути спонтанні інтерв'ю, виступи, лекції, зроблені без мікрофона-петлички. Або на самому інтерв'ю петличка може вийти з ладу, або сяде батарейка – і єдиним джерелом звуку залишиться доріжка, зроблена на накамерний мікрофон. В такому випадку реверберація буде дуже шкідливим ефектом, що значно знизить розбірливість мови і буде сильно дратувати слухачів.

Задача дереверберації мови у медіавиробництві є дуже актуальною. Метою кваліфікаційної роботи є аналіз методів дереверберації звукозапису голосу з метою визначення методів, що найменше спотворюють якість корисного сигналу і можуть бути успішно застосовані в медіаіндустрії.

В першому розділі проведено аналіз процесів реверберації в звукозаписі та їх впливу на якість сприйняття мови. Процес реверберації можна характеризувати імпульсним відгуком приміщення. Він складається з прямого звуку, ранніх відбиттів і пізніх відбиттів. Співвідношення між цими складовими залежить як від акустичних властивостей приміщення, так і від відстані від джерела до мікрофону.

Виділяють поняття критичної відстані, на якій енергія відбитих хвиль дорівнює енергії прямої хвилі. Є також поняття акустичного відношення, це відношення енергії відбитих хвиль до енергії прямої хвилі. Відповідно виділяють три звукових плани – близький, середній і віддалений.

Як показують дослідження, на якість сприйняття мови негативно впливають тільки пізні відбиття. Ранні відбиття зливаються з прямим звуком,

підсилюють його, виділяють на фоні шумів і фарбують спектр, накладаючи інформацію про приміщення. Тому показник якості в чисельнику містить прямий звук і ранні відбиття, а в знаменнику – те, що погіршує звук – пізні відбиття і шум. Отже, пізня реверберація та шум є основними основні причини погіршення розбірливості мовлення.

У другому розділі наведена класифікація методів дереверберації. можна розділити на одно-мікрофонні і мульти-мікрофонні методи. Оскільки в нашому випадку йде мова про дереверберацію в звукозаписі, тобто після процесу, в роботі розглянуто одноканальні системи. З іншого боку, способи дереверберації можна розділити на ті, що зосереджені на забарвленні мови (рання реверберація) і ті, які зосереджені на усуненні пізньої реверберації. В роботі розглянуто останні, бо саме пізня реверберація спотворює мову.

Розглянуто метод дереверберації, в якому мова синтезується генерованою послідовністю зі змінним в часі полюсним фільтром, що є поширеним способом опису мовного сигналу і називається лінійним передбаченням. Даний метод добре підходить для боротьби з реверберацією в телефонних системах зв'язку, але для медіавиробництва такий метод не підійде, бо має суттєві спотворення мовного сигналу. Тобто розбірливість зберігається, але художня якість втрачається.

Розглянуто метод спектрального віднімання. Для зменшення пізньої реверберації в спотвореному сигналі, короткочасний спектр пізньої реверберації віднімається від короткотривалого спектру вхідного сигналу з реверберацією. Фазові співвідношення при цьому залишаються тими же.

Отже, для роботи даного метода має бути оцінений коротостроковий спектр пізньої реверберації для того, щоби знайти оцінку фільтра спектрального віднімання. Чим точніше буде виконана оцінка, тим ефективніше буде дереверберація при менших спотвореннях корисного сигналу.

Розглянуто метод гомоморфної деконволюції. Цей метод базується на використанні кепстрального аналізу мовного сигналу. Сигнали, які в

поєднанні згорткою в часовій області, в комплексному кепстрі поєднанні адитивно. В результаті деконволюція зводиться до віднімання кепстрів.

Розроблено процедури, в результаті яких комплексний кепстр обчислюється із сегментів ревербераційної мови, до якої застосовується експоненційно зважена віконна функція і компоненти косинусного перетворення, з відповідною імпульсною характеристикою будуть видалені.

На базі аналізу всіх алгоритмів усунення реверберації можна зробити висновок, що найбільш придатним для подальшого використання є алгоритм спектрального віднімання. Оскільки в нашому випадку ми не маємо достатньої інформації про джерело сигналу і тільки часткові знання про канал поширення сигналу.

Існує декілька різновидів методу спектрального віднімання для дереверберації. Вони розрізняються способом оцінки імпульсної характеристики приміщення на основі реверберованого сигналу.

В методі Лембарта для того щоб отримати амплітуду короткострокового спектру пізньої реверберації, пропонується обчислити спектральну щільність потужності пізньої реверберації. Взавши припущення що енергія відбиттів більша ніж енергія прямого звуку, кореляція вхідного сигналу може бути пов'язана з кореляцією пізніх відбиттів. З цих міркувань можна вивести оцінку для спектра пізньої реверберації.

В методі Ву та Ванга пропонується виділяти пізні відбиття шляхом згладжування обвідної за допомогою функції Релея. При цьому треба мати апріорну інформацію про час реверберації і час приходу пізніх відбиттів.

В методі Фарая та Катаоки пропонується оцінювати короткочасний спектр пізньої реверберації в кожному кадрі шляхом підсумовування відфільтрованих, старих версій кадрів реверберованого сигналу. Спочатку розраховуються вагові коефіцієнти кожного кадру спектру, а потім спектр потужності попередніх кадрів, які додаються з відповідними ваговими коефіцієнтами. Вагові коефіцієнти роблять можливим вплив попереднього кадру спектру сигналу з реверберацією на спектр поточного кадру. Чим

більшу схожість показують два спектра, тим сильнішим є вплив їх попереднього спектру на поточний кадр для складової пізньої реверберації.

Для нашої роботи більш підходить метод Лембарта, оскільки він використовує час реверберації для розрахунків, що дозволить підняти точність обробки при відомих параметрах приміщення. Також це дозволяє більш детально проаналізувати якість роботи алгоритму, оскільки можна за допомогою генераторів штучної реверберації задавати час реверберації та досліджувати результати фільтрації.

Моделювання виконувалося в програмному середовищі MATLAB за таким алгоритмом: вхідний сигнал подається на генератор штучної реверберації на виході отримуємо сигнал з реверберацією; даний сигнал подається на схему усунення пізньої реверберації, після чого отримуємо відфільтрований сигнал; порівнюємо вхідний чистий сигнал та відфільтрований сигнал після роботи алгоритму за допомогою спектрограм та середньоквадратичної різниці.

Для більшої перевірки роботи алгоритму проведено моделювання на простому сигналі з декількома гармоніками (гітарному акорді), з великим часом реверберації (1,5 с). Реверберований сигнал має „хвіст“ пізньої реверберації, а відновлений сигнал має вже подавлену пізню реверберацію. Тому можна зробити висновок, що реалізація алгоритму є працездатною. Аналогічну картину можна спостерігати на спектрограмах чистого, реверберованого і відновленого сигналів.

Перевірена робота алгоритму для випадку реального мовного сигналу. На спектрограмі видно, що протяжність пізньої реверберації на різних частотах є різною. Тобто, імпульсна характеристика приміщення має чітко виражену частотну залежність. В акустиці це явище називається частотною залежністю часу реверберації. Алгоритм показав свою працездатність і на мовному сигналі.

Проведено аналіз роботи алгоритму дереверберації в залежності від параметрів реверберації. Дослідження проводились з використанням мовного

сигналу та генератора штучної реверберації Мура. Оцінювалася середньоквадратична різниця між чистим та дереверберованим сигналом. Слід зазначити, що завжди буде якийсь мінімальний рівень *RMS* оскільки в сигналі також присутня рання реверберація.

Вияснено, що *RMS* падає зі зростанням порядку моделі штучного ревербератора. В моделі Мура з порядком 5 імпульсна реакція звучить надто дискретно, що призводить до зернистості звуку, а в моделі з порядком 20 пізні відбиття більш розсіяні. Отже, метод дереверберації методом спектрального віднімання гірше працює з імпульсними звуковими сигналами і краще – з сигналами, що мають плавну атаку. Тобто, саме з мовними сигналами, метод спектрального віднімання працює краще.

При збільшенні звукопоглинання *RMS* хоч і не значно, але зростає. При зростанні звукопоглинання кількість пізніх відбиттів різко зменшується. В той же час, саме пізня реверберація дозволяє алгоритму набагато точніше оцінити імпульсну характеристику приміщення. Отже алгоритм спектрального віднімання доцільно застосовувати тільки при сильній реверберації і високому рівні пізніх відбиттів.

Залежність *RMS* від об'єму приміщення підтверджує висновок, що алгоритм спектрального віднімання ефективніше працює при тривалій реверберації і високому рівні пізніх відбиттів. Тому що при збільшенні об'єму приміщення пізні відбиття стають більш вираженими, що дозволяє алгоритму точніше визначити імпульсну характеристику приміщення.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Naylor P., Gaubitch N. Speech Dereverberation. – Springer, 2010. – 399 p.
2. B.D. Radlović, R. Williamson, R. Kennedy. On the poor robustness of sound equalization in reverberant environments // Proc. of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'99), 1999, vol. 2. – pp. 881–884.
3. B.D. Radlović, R. Williamson, R. Kennedy. Equalization in an acoustic reverberant environment: robustness results // IEEE Trans. Speech Audio Processing, vol. 8, no. 3, 2000. – pp. 311–319,
4. F. Talantzis and D. Ward. Investigation of performance of acoustic arrays for equalization in a reverberant environment // Proc. of the fourth International Conference on Digital Signal Processing, 2002, vol. 1. – pp. 247–250.
5. J.S. Garofolo. Getting started with the darpa timit cd-rom: An acoustic phonetic continuous speech database // Tech. rep., National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, Dec. 1988.
6. X. Huang, A. Acero, H. Hon. Spoken language processing. A guide to theory, algorithm and system development, Prentice-Hall, 2001.
8. G.A. Soulodre, N. Popplewell, J.S. Bradley. Combined effects of early reflections and background noise on speech intelligibility // Journal of Sound Vibration, vol. 135, 1989. – pp. 123–133.
9. J.C. Steinberg. Effects of distortion upon the recognition of speech sounds // Journal of the Acoustical Society of America, vol. 1, 1995. – pp. 121–137.
10. M. Hodgson and E.M. Nosal. Effect of noise and occupancy on optimal reverberation times for speech intelligibility in classrooms // Journal of the Acoustical Society of America, vol. 111, 2002. – pp. 931–939.
11. J.P.A. Lochner and J.F. Burger. The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech // Acustica, vol. 8, 1958. – pp. 1–10.

12. A.J. Watkins and N.J. Holt, “Effects of a complex reflection on vowel identification // *Acustica*, vol. 86, 2000. – pp. 532–542.

13. F. Aigner and M.J.O. Strutt. On a physiological effect of several sources of sound on the ear and its consequences in architectural acoustics // *Journal of the Acoustical Society of America*, vol. 6, no. 3, 1935. – pp. 155–159.

14. V.M.A. Peutz. Articulation loss of consonants as a criterion for speech transmission in a room // *Journal of the Audio Engineering Society*, vol. 19, no. 11, 1971. – pp. 915–919.

15. D.A. Berkley. Acoustical factors affecting hearing aid performance, chapter Normal listeners in typical Rooms – Reverberation Perception, Simulation, and Reduction. University Park Press, Baltimore, 1980. – pp. 3–24.

16. J.B. Allen. Effects of small room reverberation on subjective preference // *Journal of the Acoustical Society of America*, vol. 71, no. S1, 1982. pp. 50-55.

17. J.J. Jetzt. Critical distance measurement of rooms from the sound energy spectral response // *Journal of the Acoustical Society of America*, vol. 65, no. 5, 1979. – pp. 1204–1211.

18. В.Н. Олейников, О.В. Зубков, В.М. Карташов, И.В. Корытцев, С.И. Бабкин, С.А. Шейко, И.С. Селезнев. Экспериментальная оценка эффективности алгоритмов пеленгования беспилотных летательных аппаратов по акустическому излучению. *Радиотехника: Всеукр. межвед. науч.-техн. сб.* – 2019. – Вып. 199. – С. 29 – 37.

19. 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 Confer-ence on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET). 2020. 4 p.

20. Kartashov V.M., Oleynikov V.N, Zubkov O.V., Korytsev 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.

21. 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.

22. Буйницький Д.В. Апаратно-программний комплекс для дослідження акустических пристроїв // 23-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті». – Харків, 16–18 квітня 2019 р. – с. 92-93.

23. Ашихмин В.О. Исследование системы коррекции звука с учётом характеристик помещения // 23-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті». – Харків, 16–18 квітня 2019 р. – с. 96-97.

24. Тушев В.О. Исследование влияния фазовых искажений аудиоаппаратуры на качество звучания // 23-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті». – Харків, 16–18 квітня 2019 р. – с. 84-85.

25. Чернов К.А. Исследование акустических характеристик студии звукозаписи // 23-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у ХХІ столітті». – Харків, 16–18 квітня 2019 р. – с. 94-95.

26. Ашихмін В.О. Адаптивні методи подавлення акустичного зворотного зв'язку: атестаційна робота на здобуття освітнього ступеня магістра. – Х.: ХНУРЕ. – 2019 р. – 69 с.

28. Буйницький Д.В. Дослідження методів компенсації ревербераційних завад: атестаційна робота на здобуття освітнього ступеня магістра. – Х.: ХНУРЕ. – 2019 р. – 67 с.

29. Шевченко П.Д. Система вимірювання та аналізу акустичних характеристик приміщень: атестаційна робота на здобуття освітнього ступеня магістра. – Х.: ХНУРЕ. – 2019 р. – 68 с.

30. Лісовий Р.О. Комплексне дослідження акустичних характеристик студійного приміщення: атестаційна робота на здобуття освітнього ступеня магістра. – Х.: ХНУРЕ. – 2019 р. – 81 с.

31. Древальський Р.В. Дослідження методу корекції звука для компенсації впливу приміщення /25-й Міжнародний молодіжний форум «Радіоелектроніка та молодь у XXI столітті». Зб. матеріалів форуму. Т. 3. – Харків: ХНУРЕ. 2021. – с. 119 – 120.

32. E.A.P Habets. Single- and multi-microphone speech dereverberation using spectral enhancement. Ph.D. dissertation, T. U. Eindhoven, 2007.

33. Hardwick, C.D. Yoo, and J.S. Lim. Speech enhancement using the dual excitation speech model // in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 1993. – pp. 367–370.

34. C.D. Yoo. Speech enhancement based on the generalized dual excitation model with adaptive analysis window // in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 1995, vol. 1. – pp. 832–835.

35. H. Attias and L. Deng. Speech Denoising and Dereverberation Using Probabilistic Models // Advances in Neural Inform. Process. Syst., vol. 13, 2001. – pp. 758–764.

36. J.R. Deller, J.G. Proakis, and J.H.L. Hansen. Discrete-Time Processing of Speech Signals. New York: MacMillan, 1993.

37. M. Tonelli, M.G. Jafari, and M.E. Davies. A multi-channel Maximum Likelihood approach to de-reverberation // in Proc. European Signal Process. Conf. (EUSIPCO), Florence, Italy, Sept. 2006.

38. B. Yegnanarayana. Speech enhancement using excitation source information // in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), 2002, vol. 1. – pp. 541–544.

39. H. Wang and F. Itakura. An implementation of multi-microphone dereverberation approach as a preprocessor to the word recognition system // J. Acoust. Soc. Jap., vol. 13, no. 5, 1992. – pp. 285–293.

40. A. Oppenheim, R. Schafer, T. Stockham. Nonlinear filtering of multiplied and convolved signals // *IEEE Trans. Audio and Electroacoustics*, Vol. AU-16, No. 3, 1968.

41. M. Wu and D. L. Wang. A two-stage algorithm for one microphone reverberant speech enhancement // *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 3, 2006. – pp. 774–784.

42. P. Krishnamoorthy and S. R. M. Prasanna. Reverberant speech enhancement by temporal and spectral processing // *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 2, 2009. – pp. 137–148.

43. S. Mosayyebpour, M. Esmaeili, T. A. Gulliver. Single-Microphone Early and Late Reverberation Suppression in Noisy Speech // *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 2, Feb. 2013. – pp. 322–335.

44. E. A. P. Habets, N. Gaubitch, and P. A. Naylor. Temporal selective dereverberation of noisy speech using one microphone // in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2008. – pp. 4577–4580.

45. J. S. Erkelens and R. Heusdens. Single-microphone late-reverberation suppression in noisy speech by exploiting long-term correlation in the DFT domain // in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2009. – pp. 3997–4000.

46. 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.