

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 алгоритмів автоматизованої еквалізації звукозапису голосу.
(тема)

Виконав:
студент 2 курсу, групи МІМ-21-1
Мезенцев І.О.
(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. Цехмістро Р.І.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри Володимир КАРТАШОВ
(підпис)

2022 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

студентові Мезенцеву Івану Олексійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження алгоритмів автоматизованої еквалізації звукозапису голосу.

затверджена наказом по університету від " 24 " 10 2022 р. № 1384 Ст _____

2. Термін подання студентом роботи до екзаменаційної комісії 07.12.2022 р.

3. Вихідні дані до роботи Виконати теоретичні оцінки факторів, що впливають на наскрізну АЧХ тракту формування, обробки, передачі і відтворення мовного сигналу. Визначити головні фактори, що впливають на якісні показники (натуральність) звучання мови диктора. Отримати вираз для оціночної кривої АЧХ еквалайзера корекції. Отримати теоретичні оцінки характерних особливостей спектрів мови диктора і швидкості спадання спектральної щільності потужності. Провести уточнення теоретичних результатів шляхом експериментального дослідження аудіо записів реальних ефірів на радіо, телебаченні і в Ютубі. Провести експериментальне дослідження корекції АЧХ мови диктора для оцінки можливостей і обмежень автоматизованої системи корекції.

4. Перелік питань, що потрібно опрацювати в роботі _____

Вступ

1. Аналітичний огляд основних параметрів та характеристик мовних сигналів, а також методів їх частотної корекції.

2. Теоретичне обґрунтування принципів еквалізації мовних сигналів.

3. Експериментальне дослідження спектрів мовних сигналів і їх автоматичної еквалізації.

Висновки

Перелік посилань

Додатки

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

1. Еквалізація голосу (1 аркуш А4).
2. Постановка задачі (1 аркуш А4).
3. Структурна схема тракту запису (1 аркуш А4).
4. Оцінка АЧХ тракту звукозапису (1 аркуш А4).
5. Теоретичні оцінки спектру мови (1 аркуш А4).
6. Експериментальне дослідження записів ефіру (1 аркуш А4).
7. Результати розрахунку нахилів спектрів (1 аркуш А4).
8. Експериментальна обробка записів мови (1 аркуш А4).
9. Висновки (1 аркуш А4).

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналітичний огляд літератури	01.09.22–13.09.22	
2	Обґрунтування принципів еквалізації	14.09.22–27.09.22	
3	Експериментальне дослідження спектрів	28.09.22–11.10.22	
4	Експериментальне дослідження еквалізації	12.10.22–25.10.22	
5	Обробка результатів	26.10.22–10.11.22	
6	Графічна частина роботи	11.11.22–25.11.22	
7	Перевірка керівником	26.11.22–02.12.22	
8	Перевірка на академічний плагіат	03.12.22	
9	Перевірка завідувачем кафедри, рецензування	04.12.22–07.12.22	

Дата видачі завдання _____ 01.09.2022 р. _____

Студент _____ Іван МЕЗЕНЦЕВ _____
(підпис)

Керівник роботи _____ Роман ЦЕХМІСТРО _____
(підпис)

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 69 сторінок, 39 рисунків, 2 таблиці, 33 джерела.

АЧХ, ЕКВАЛІЗАЦІЯ, КОРЕКЦІЯ, МОВНИЙ СИГНАЛ, СПЕКТР, ТЕМБР, ЧАСТОТА, ШВИДКЕ ПЕРЕТВОРЕННЯ ФУР'Є

Об'єкт дослідження – процеси спотворення і корекції частотних характеристик звукозаписів мови диктора.

Мета роботи – розробити методику оцінки нерівномірності АЧХ в тракті звукозапису без використання спеціальних випробувальних сигналів з подальшим виправленням цих нерівномірностей коригуючим еквайзером; визначення можливостей і обмежень автоматизованої системи корекції.

В роботі виконано теоретичну оцінку факторів, що впливають на наскрізну АЧХ тракту формування, обробки, передачі і відтворення мовного сигналу. Визначено головні фактори, що впливають на якісні показники (натуральність) звучання мови диктора. Отримано вираз для оціночної кривої АЧХ еквайзера корекції. Отримано теоретичні оцінки характерних особливостей спектрів мови диктора і швидкості спадання спектральної щільності потужності. Проведено уточнення теоретичних результатів шляхом експериментального дослідження аудіозаписів реальних ефірів на радіо, телебаченні і в Тубі. Проведено експериментальне дослідження корекції АЧХ мови диктора для оцінки можливостей і обмежень автоматизованої системи корекції.

ABSTRACT

Explanatory note to the qualification work: 69 pages, 39 figures, 2 tables, 33 sources.

AFFECTIVENESS, EQUALIZATION, CORRECTION, SPEECH SIGNAL, SPECTRUM, TIMBRE, FREQUENCY, FAST FOURIER TRANSFORM

The object of research is the processes of distortion and correction of frequency characteristics of sound recordings of the announcer's speech.

The purpose of the work is to develop a methodology for assessing frequency response unevenness in the sound recording path without using special test signals with subsequent correction of these unevennesses with a corrective equalizer; determination of capabilities and limitations of the automated correction system.

In the paper, a theoretical assessment of the factors affecting the end-to-end frequency response of the formation, processing, transmission and reproduction of the speech signal was performed. The main factors affecting the quality indicators (naturalness) of the announcer's speech are determined. The expression for the estimated frequency response curve of the correction equalizer is obtained. Theoretical estimates of the characteristic features of the speaker's speech spectra and the rate of decline of the power spectral density were obtained. The theoretical results were refined through experimental research of audio recordings of real broadcasts on radio, television and Tube. An experimental study of the frequency response correction of the announcer's speech was conducted to assess the capabilities and limitations of the automated correction system.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів.....	8
Вступ.....	9
1 АНАЛІТИЧНИЙ ОГЛЯД ОСНОВНИХ ПАРАМЕТРІВ ТА ХАРАКТЕРИСТИК МОВНИХ СИГНАЛІВ, А ТАКОЖ МЕТОДІВ ЇХ ЧАСТОТНОЇ КОРЕКЦІЇ.....	11
1.1 Основні параметри та характеристики мовних сигналів.....	11
1.2 Основні інтегральні параметри мовних сигналів.....	13
1.3 Частотні характеристики голосу.....	14
1.4 Тембральні характеристики голосу.....	15
1.5 Частотна корекція звуку.....	18
1.5.1 Еквалайзери.....	18
1.5.2 Частотні діапазони звукових сигналів.....	21
1.5.3 Амплітудно-частотні характеристики фільтрів.....	23
1.5.4 Фазо-частотні характеристикти фільтрів.....	24
1.5.5 Імпульсний відгук фільтрів.....	26
1.6 Висновки по розділу 1.....	27
2 ТЕОРЕТИЧНЕ ОБГРУНТУВАННЯ ПРИНЦИПІВ ЕКВАЛІЗАЦІЇ МОВНИХ СИГНАЛІВ.....	29
2.1 Тракт формування, обробки, передачі і відтворення мовного сигналу.....	29
2.2 Тонка структура спектра і середня спектральна щільність потужності.....	33
2.3 Врахування властивостей слуху.....	38
2.4 Висновки за розділом 2.....	41
3 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ СПЕКТРІВ МОВНИХ СИГНАЛІВ І ЇХ АВТОМАТИЧНОЇ ЕКВАЛІЗАЦІЇ.....	43

3.1 Обґрунтування методу вимірювання спектру мовних сигналів.....	43
3.2 Дослідження спектрів мовних сигналів.....	50
3.3 Експериментальне дослідження еквалізації мовних сигналів.....	55
3.4 Висновки по розділу 3.....	61
Висновки.....	63
Перелік джерел посилань.....	66
ДОДАТКИ.....	70
Додаток А. Графічний матеріал.....	71
Додаток Б. Відомість кваліфікаційної роботи.....	80

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,
СКОРОЧЕНЬ І ТЕРМІНІВ

- АЧХ – амплітудно-частотна характеристика;
ВЧ – верхні частоти;
Г – гучномовець;
ГЧЗ – груповий час запізнення;
дБ – децибели;
ДПФ – дискретне перетворення Фур'є;
ЗвВ – звуковідтворювальна апаратура;
ЗвЗ – звукозаписувальна апаратура;
КЗ – канал зв'язку;
М – мікрофон;
МС – мовний сигнал;
НЧ – нижні частоти;
О – октавність еквайзера;
ОТ – основний тон;
Пр – приміщення студії;
СМ – сигнал маскування;
СЩП – спектральна щільність потужності;
ТПМ – тракт передачі мови;
Ф – форманта мови;
ФЧХ – фазо-частотна характеристика;
ANSI – American national standards institute – Американський національний інститут стандартів;
Frequency – центральна (робоча) частота;
Gain – підсилення чи ослаблення обраної смуги;
Q – добротність;
SPL – Sound Pressure Level – рівень звукового тиску.

ВСТУП

Запис голосу диктора – дуже важливий етап аудіовиробництва. Але запис – це лише половина справи. У даній кваліфікаційній роботі розглядається процес обробки голосу диктора.

Після того, як запис диктора виконано, настає етап редагування. Людський голос надзвичайно виразний сам по собі, проте за допомогою різних методів обробки звуку можна отримати ще цікавіші та яскравіші голосові ефекти. Для голосу можна застосувати весь спектр існуючих методів обробки – амплітудних, частотних, фазових, часових тощо.

Частотна корекція або як кажуть звукорежисери – еквалізація – один з найпопулярніших ефектів для обробки голосу. Також, на жаль, це один із ефектів, яким часто зловживають. При неправильній еквалізації, кінцевий результат може звучати навіть гірше, ніж до використання.

Звичайно ж, не існує якихось суворих правил, згідно з якими має оброблятися частотний спектр голосу, але є основні принципи, про які й йтиметься в роботі.

Об'єкт дослідження – процеси спотворення і корекції частотних характеристик звукозаписів мови диктора.

Мета роботи – розробити методику оцінки нерівномірності АЧХ в тракту звукозапису без використання спеціальних випробувальних сигналів з подальшим виправленням цих нерівномірностей коригуючим еквалайзером; визначення можливостей і обмежень автоматизованої системи корекції.

Задачі дослідження:

- виконати теоретичну оцінку факторів, що впливають на наскрізну АЧХ тракту формування, обробки, передачі і відтворення мовного сигналу;
- визначити головні фактори, що впливають на якісні показники (натуральність) звучання мови диктора;
- отримати вираз для оціночної кривої АЧХ еквалайзера корекції;

- отримати теоретичні оцінки характерних особливостей спектрів мови диктора і швидкості спадання спектральної щільності потужності;
- провести уточнення теоретичних результатів шляхом експериментального дослідження аудіо записів реальних ефірів на радіо, телебаченні і в Ютубі;
- провести експериментальне дослідження корекції АЧХ мови диктора для оцінки можливостей і обмежень автоматизованої системи корекції.

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

1 АНАЛІТИЧНИЙ ОГЛЯД ОСНОВНИХ ПАРАМЕТРІВ ТА ХАРАКТЕРИСТИК МОВНИХ СИГНАЛІВ, А ТАКОЖ МЕТОДІВ ЇХ ЧАСТОТНОЇ КОРЕКЦІЇ

1.1 Основні параметри та характеристики мовних сигналів

Мова як процес генерації мовних сигналів на загальну думку [1,2] є квазішумовим квазістаціонарним сигналом з нестійкими параметрами та характеристиками складових її елементів. У різних словах одні й ті самі звуки можуть вимовлятися по-різному, а частотні характеристики мовоутворюючого тракту змінюються залежно від віку та психологічного стану людини. У даному випадку мовний сигнал можна аналізувати як шумоподібний акустичний сигнал [3], що має інформаційну амплітудну (параметр A_i), частотну (параметр ω_i) та фазову (параметр φ_i) модуляцію несійних n елементарних гармонійних сигналів:

$$S(t) = \sum_{i=1}^n S_i(t) = \sum_{i=1}^n A_i \cos[\omega_i t + \varphi_i(t)]. \quad (1.1)$$

Джерелами генерації інформаційних акустичних гармонійних елементарних сигналів $S_i(t)$ у мовотворному тракті є:

- періодична модуляція повітряного потоку, що виходить з легень, за допомогою колихаючоїся діафрагми та голосових зв'язок (голосове джерело сигналів основного тону (ОТ));
- турбулентні завихрення повітряного потоку у звуженнях мовотворного тракту (шумове джерело);
- стрибкоподібна зміна тиску повітря в мовотворному тракті при різкому короткому перериванні закінчення повітря (імпульсне джерело).

Вважається [4], що елементарні гармонічні сигнали артикуюються (модулюються) за амплітудою та фазою в акустичному фільтрі, який

утворюють активні та пасивні артикуляційні органи та система акустичних резонаторів. Внаслідок його роботи окремі складові мовного сигналу посилюються чи послаблюються.

Під терміном форманта (Φ_n , n – номер форманти) розуміється [6] гармонічні компоненти сигналів основного тону (ОТ), виділені та підсилені в акустичному фільтрі мовотворного тракту.

Розглянемо основні диференціальні параметри та характеристики мовного сигналу на рівні формант. Наведені нижче параметри та характеристики в основному визначаються статистичною обробкою [7] квазистаціонарної ділянки мовного сигналу, яка визначається як складовий інтервал $T_c \approx 10 \dots 20$ мс.

Це насамперед спектральна щільність потужності (СЩП), статистичні характеристики Φ_n та сигналу ОТ.

Частотне розподілення формант Φ_n наступне:

- форманти дзвінких звуків займають смуги частот (150 – 900) Гц (Φ_1), (550 – 2800) Гц (Φ_2) та (1500 – 3400) Гц (Φ_3);
- форманти глухих звуків займають смуги частот (1000 – 3500) Гц (Φ_1), (2500 – 6000) Гц (Φ_2) та (1500 – 4000) Гц (Φ_3);
- з ймовірністю 0,98 частотний діапазон Φ_1 дорівнює (200 – 850) Гц, Φ_2 дорівнює (850 – 2550) Гц та Φ_3 дорівнює (2100 - 3300) Гц.

Девіація медіанної частоти формант характеризується коефіцієнтом взаємної кореляції, що дорівнює 0,78 для Φ_1 та Φ_2 , 0,82 для Φ_2 та Φ_3 , і 0,95 для Φ_3 та Φ_4 . Середнє значення девіації частоти обвідної для формантів Φ_1 і Φ_2 порядку 8 Гц [3].

Амплітуда сигналу визначає гучність мовних сигналів. Частота коливань голосових зв'язок, характеризує сигнал ОТ, становить

- від 70 до 180 Гц (середня частота 129 Гц) для чоловічих голосів та
- від 180 до 330 Гц (середня частота 240 Гц) для жіночих.

Процес модуляції сигналів ОТ акустичним фільтром мовотворного тракту визначають як вокалізацію елементів мовного сигналу. В одного і того

ж диктора залежно від емоційного стану та ситуативності мови частота сигналу ОТ може змінюватись.

Інформаційні параметри та характеристики зміни індексу фазової модуляції (за параметром $\phi_i(t)$ виразу (1.1)) статистично не визначені [3].

1.2 Основні інтегральні параметри мовних сигналів

Основну енергію акустичних коливань МС укладено у діапазоні (70 – 7000) Гц, причому понад 95% семантичної інформації розміщується у вузкому частотному діапазоні (200 – 5000) Гц.

Акустичні коливання вище та нижче цього діапазону частот несуть ідентифікаційну інформацію про емоції та особистості диктора (усний почерк), яка сприяє впізнаваності, і дещо підвищують розбірливість мовних сигналів за умов інтенсивних природних маскуючих сигналів.

Динамічні показники розмовної мови різні і багато в чому залежать від зовнішніх умов, у яких знаходиться диктор. Так:

- спокійна, довірча розмова, що ведеться співрозмовниками, які перебувають поруч один з одним відбувається зазвичай з рівнем порядку 55 дБ SPL (звуковий тиск);

- виступ у залі, а часто й розмова по телефону – близько 75 дБ SPL.

При цьому динамічний діапазон мовного сигналу також змінюється у досить широких межах 25 – 45 дБ.

Розподіл амплітудного складу мовного сигналу показує [5], що понад 80% звуків мовлення мають рівень менше 45 дБ, і, вважається, легко можуть бути приховані сигналом маскування (СМ). Локалізовані елементи мовного сигналу мають основну несійну частоту в межах (80 – 2500) Гц та тривалість $T_c \approx 30 \dots 300$ мс, у яких зосереджена основна енергія мовного сигналу.

Відповідно до методу наукової абстракції для аналізованого предмета досліджень інформаційну складову сигналу $S(t)$ можна подати як суму вокалізованих частотних компонент, з певними межами зміни параметрів

сигналу ОТ і сигналів перших трьох формант. Представляючи сигнал ОТ як нульову форманту (Φ_0) абстрагуємо вираз (1.1) для інформаційних (у частині семантичної адекватності) компонент сигналів $S(t)$

$$S(t) = \sum_{i=1}^3 A_{\Phi_i} \cos[\omega_{\Phi_i} t + \varphi_{\Phi_i}(t)]. \quad (1.2)$$

Параметр адекватності прийнятого мовного інформаційного сигналу оцінюється розбірливістю, яка є статистичною характеристикою мовного сигналу $S(t)$ прийнятої на фоні перешкоди $S_n(t)$.

Під терміном розбірливість розуміють [6] відношення числа правильно прийнятих елементів мовного сигналу (складів, слів, фраз із словника обсягом V_w) до загального числа елементів, переданих диктором. В якості показника оцінки завадозахищеності найчастіше використовують словесну розбірливість W , яка визначає семантичну адекватність інформації та досить об'єктивно оцінює інформаційну ефективність передавання мовного сигналу.

1.3 Частотні характеристики голосу

Спектральні характеристики всіх звукових сигналів у загальному випадку подібні і описані в більшості книг курсу загальної фізики.

Визначаються вони за допомогою двох основних залежностей, які розкривають їх склад повною мірою:

– залежність амплітуди від часу (осцилограма або амплітудно-часова характеристика) та

– залежність амплітуди від частоти (спектрограма чи амплітудно-частотна характеристика).

Особливістю мовних сигналів є наявність формантних областей – набору гармонік, що утворюють певний звук мови [6].

Всі мовні сигнали можна умовно поділити на те, що схоже у певного складу людей – це відношення до будь-якої мовної групи: подібний набір слів і звуків, що вимовляються, і те, що є індивідуальним для кожної людини – низка унікальних особливостей, виражених у фізичній індивідуальності мови, вимові тембру голосу, так званому артикуляційному апараті людини.

При формуванні переліку спектральних характеристик з формантних областей береться до уваги лише перша (основна) гармоніка, оскільки вона є найінформативнішою. Інші гармоніки лише визначають сказаний звук, що здебільшого є загальним у людей одного мовного складу.

Відповідно до вимірюваної величини основні спектральні показники розділені на такі підгрупи: частотні, енергетичні та часові. Мовний сигнал є не постійною величиною та для повної картини його складу необхідний контроль значень у кожний момент часу, тому для кожної характеристики визначається її мінімальне, середнє та максимальне значення.

Важливим уточненням є те, що мовний сигнал є безперервним у часі, і під гармонікою розуміється деяка спектральна щільність із вираженим піком амплітуди. Частотні характеристики мови:

- частота основного тону – F_{OT} (Гц) – частота першої гармоніки спектру;
- період основного тону – $T_{OT} = 1/F_{OT}$ (мс);
- кількість побічних гармонік (формантних областей) – N_2 ;
- нижня частота спектра – $f_n = F_{OT}$ (Гц);
- верхня частота діапазону – f_e (Гц);
- частота максимального рівня спектральної густини – F_{Imax} (Гц).

1.4 Тембральні характеристики голосу

Тембр доповнює характеристику музичних звуків за висотою, гучністю, локалізації та дозволяє розрізнити між собою звуки різної природи (наприклад, співочі голоси, музичні інструменти).

В оцінці тембру мають значення кількість, розташування та інтенсивність компонентів спектру (так звані складові спектру): наприклад наявність «стійких груп складових, які називаються формантами». Якщо в звуку, то він сприймається як порожній, незабарвлений; якщо сильно виражені перші обертони, то він соковитий, повний. У разі, якщо у спектрі звуку сильно виражені вищі гармоніки (в області частот 3-6 кГц), то тембр оцінюється як різкий, "металевий".

У характеристику тембру входять також шумові призвуки, атака (початковий момент звуковидобування), вібрато та інші фактори. На тембр надають вплив параметри самого джерела звуку: матеріал, форма джерела, умови його коливальності, характеристика резонатора (резонаторів), акустика приміщення (якщо є). Слід зазначити, що науково обґрунтована типологія тембру нині відсутня.

Обертон (від нім. Oberton, буквально – верхній тон, верхній звук) називають звук, що входить у спектр музичного звуку, який звучить вище основного тону [7]. Наявність обертонів обумовлено складним коливанням тіла, що звучить, яке є накладенням (суперпозиція) простих коливань. Обертони разом із основним тоном називають частковими тонами: 1-й частковий тон дорівнює основному тону, 2-й - першому обертону і т.д.

Якщо частоти послідовних обертонів кратні частоті основного тону (утворюють разом з ним натуральний звукоряд або його початкова ділянка), то такі обертони називаються гармонійними (інакше негармонічними), а спектр звуку – гармонійними. Гармонійні обертони разом з основним тоном називають гармоніками: 1-я гармоніка відповідає основному тону, 2-я гармоніка – першому обертону і т.д. Основний тон умовно вважають нульовим обертоном. Гармоніки, починаючи з 2-ї, називають вищими.

У стандарті ANSI S1.1-1960 «Акустична термінологія» Американського національного інституту стандартів тембру дано таке визначення: «це атрибут слухового сприйняття, який дозволяє слухачеві визначити, що два звуки мають однакову висоту і гучність, відрізняються

один від одного». У 1976 р. до цього визначення у стандарті зроблено таке додавання: «тембр залежить насамперед від частотного спектру, хоча він також залежить від звукового тиску та часових характеристик звуку» [8]. Деякі дослідники [9, 10] вважають це визначення неповним і навіть невдалим.

У мовознавстві [11] поняття «тембр» розкривається так. По-перше, тембр – це якість звуку мови, що залежить від співвідношення за висотою та силою основного тону та додаткового (обертону). Завдяки тембру розрізняють голосні та сонанти (сонорні приголосні). Тембр (або забарвлення) приголосних – це їхня додаткова артикуляційно-акустична характеристика. Наприклад, носовий тембр, що купується приголосним при назалізації; м'якість чи твердість звуків. По-друге, тембр – це один із складових елементів інтонації – звукове забарвлення, що надає мовленню ті чи інші експресивно-емоційні відтінки. Наприклад, похмурий, веселий, грайливий, іронічний тембр.

На рис.1.1 наведена класифікація тембру залежно від розташування «спектрального центрoїду» у спектрі музичних тонів.

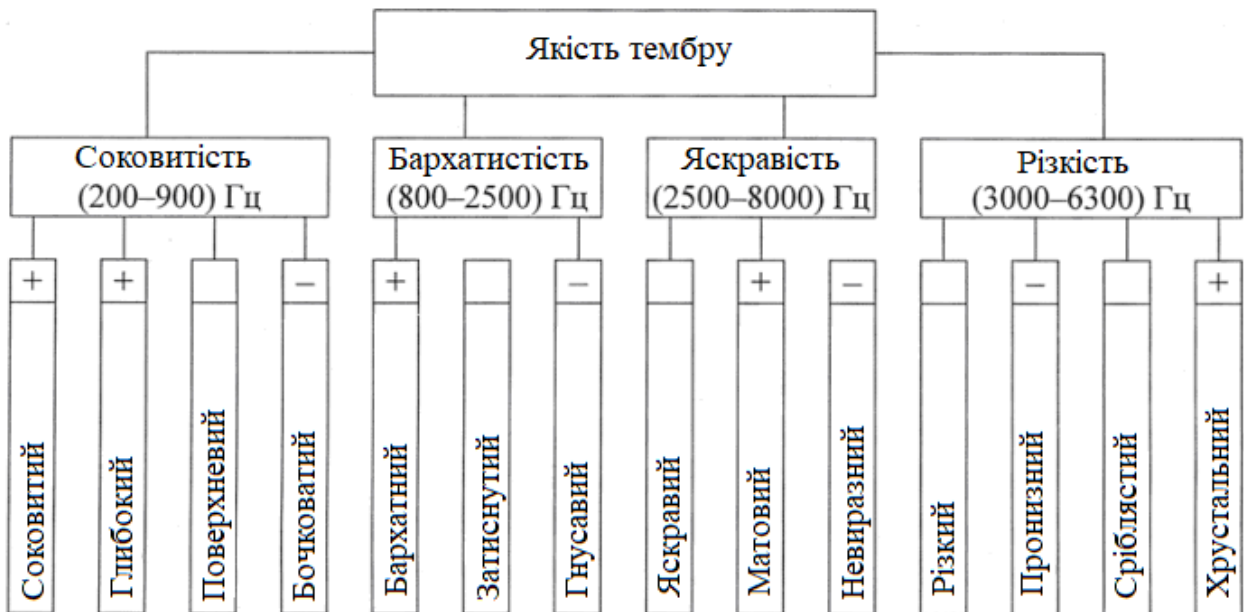


Рисунок 1.1 – Класифікація тембру залежно від розташування «спектрального центрoїду» у спектрі музичних тонів

Значимість спектральної обвідної звуку в сприйнятті тембру підтверджена дослідженнями М. Clarkson [12]. Дослідження показали, що зміни спектрального складу звуку насамперед впливають на сприйняття тембру.

При розгляді форми спектральної обвідної – розподілу звукової енергії між спектральними смугами – слід зазначити, що на оцінку суб'єктивної якості тембру впливає «спектральний центроїд» (інша назва – «центр мас»). Під ним розуміють «середню точку розподілу спектральної енергії звуку» [13], яке корелює з уявною «яскравістю» звуку. Зміщення "середнього центроїду" в область високих частот (тобто збільшення енергії групи обертонів – гармонік основного тону) відчувається слухачем як підвищення яскравості тембру звучання [13].

На тембр, що формується, впливає і угруповання обертонів у так звані «формантні групи, особливо в області максимальної чутливості слуху» [13], яка, крім цього, визначає фонемну розрізненість голосних звуків мови. З погляду І.А. Алдошиной і Р. Прітте [13], думка, що «сприймаємий тембр звуку залежить від його спектрального складу, тобто, від розташування обертонів на частотній шкалі, від співвідношення їх амплітуд, форми спектральної обвідної, спектрального розподілу енергії за частотою та ін., безумовно, справедливо та підтверджується численною практикою роботи зі звуком».

1.5 Частотна корекція звуку

1.5.1 Еквалайзери

Процес частотної обробки звукових сигналів називають "еквалізацією" (Equalization). Цей термін звичайно має на увазі корекцію або зміну відношень амплітуд сигналів з різними частотами.

Еквалайзер – це певний набір фільтрів різного типу. Є два основні типи багатосмугових еквалайзерів: графічні і параметричні.

Графічні еквайзери мають певну кількість регульованих по рівню частотних смуг, кожна з яких має постійну робочу частоту, фіксовану ширину смуги навколо робочої частоти, а також діапазон регулювання рівня (він однаковий для усіх смуг). Найчастіше, крайні смуги (найнижча і найвища) є фільтрами «полочного» типу, а всі інші мають характеристику типу «дзвону».

Октавність еквайзера O розраховують по відношенню сусідніх частот регулювання [14]:

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

де n – порядковий номер смуги регулювання еквайзера.

Графічні еквайзери, які використовуються у професійних галузях, зазвичай 1/2- чи 1/3-октавні і мають по 15 або 31 смугі на канал, часто комплектуються аналізаторами для зручності регулювання [14].

На рис.1.2 наведена АЧХ і управління 1/2-октавного еквайзера.

Параметричний еквайзер за меншої кількості налаштувань дає набагато ширші можливості регулювання частотної характеристики сигналу.

Кожна його смуга містить три регульовані параметри:

– Frequency – центральна (робоча) частота (Гц);

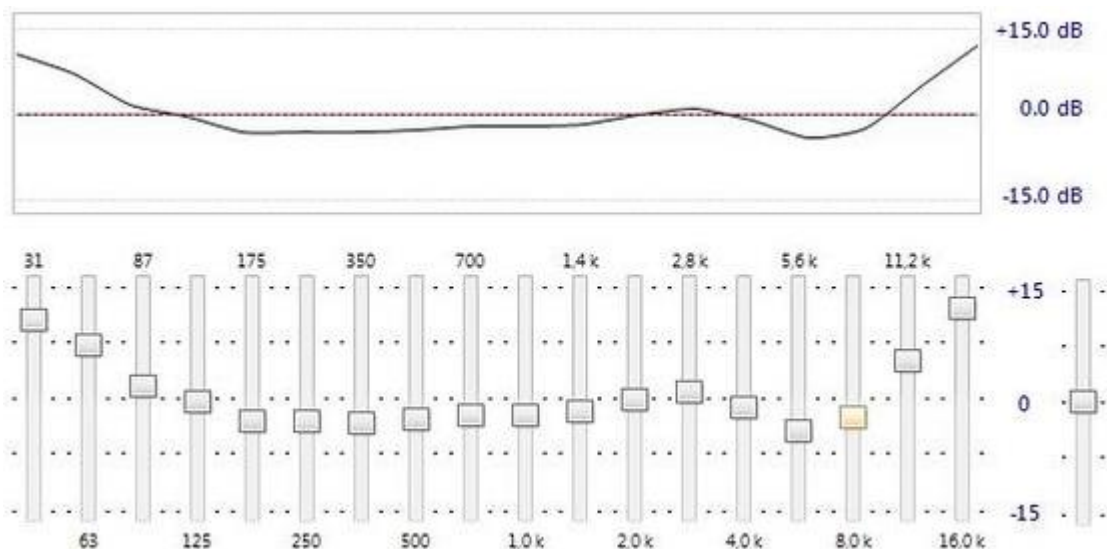


Рисунок 1.2 – Панель управління 1/2-октавного еквайзера

– Q – добротність (налаштовує ширину Δf робочої смуги частот навколо центральної частоти f_0) [14]:

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

– Gain – підсилення чи ослаблення обраної смуги (дБ).

Отже, при застосуванні параметричного еквайзера можна набагато точніше підібрати необхідну частоту і більш точно відрегулювати смугу і коефіцієнт передачі. На рис.1.3 наведена панель керування параметричного еквайзера, а на рис.1.4 – його АЧХ при граничних положеннях налаштувань Gain, Frequency, Q (добротність).



Рисунок 1.3 – Панель керування параметричного еквайзера

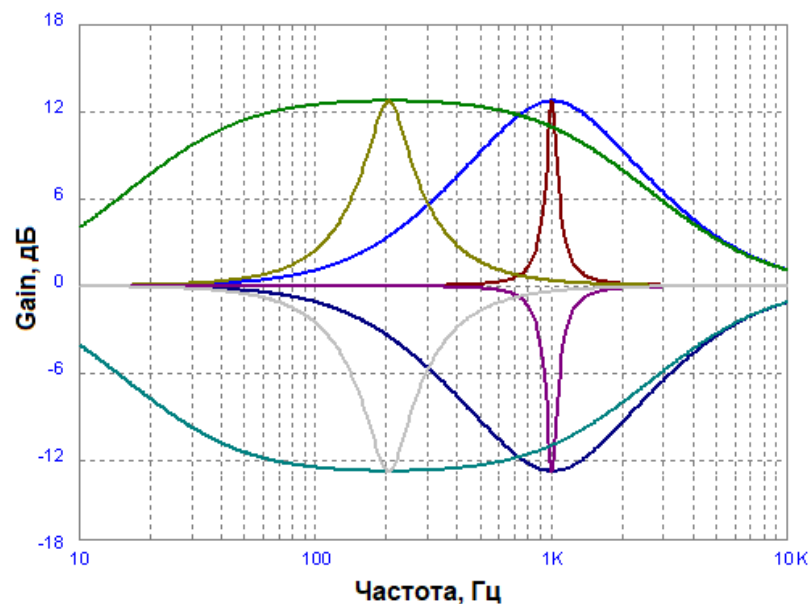


Рисунок 1.4 – АЧХ параметричного еквайзера при крайніх положеннях регулювань Gain, Frequency, Q (добротність)

Аналогові параметричні еквайзери застосовують досить нечасто, вони мають невелику кількість регульованих смуг частот. Але досягнення в цифровій обробці звукових сигналів сприяли появі цифрових параметричних еквайзерів з фактично необмеженим числом регульованих смуг частот. Крім того, в цифрових параметричних еквайзерах часто є додаткові параметри, такі як тип фільтру, характер кривої АЧХ і т.д. Дуже часто параметричні еквайзери можуть виконувати роль одного з блоків обробки в цифрових акустичних процесорах [14].

Є еквайзери змішаного типу, що можна зустріти в мікшерних консолях, де наприклад складові НЧ і ВЧ регулюються по типу графічного еквайзера «поличного» типу, а між ними знаходяться 2 напівпараметричні смуги (без налаштування добротності).

Також є гібриди, які називають параграфічними еквайзерами – це гібрид параметричного і графічного еквайзерів. Вони, як правило, дозволяють управляти величиною підсилення (gain) за допомогою повзунків чи на дисплеї, але при тому мають регулювання добротності та регулювання центральної частоти для кожної смуги.

Є просторові еквайзери, що дозволяють окремо коригувати моно (mid) і стерео (side) складові звукового сигналу, а також стереоеквайзери, в яких є можливість еквалізувати окремо правий і лівий канали стереосигналів.

1.5.2 Частотні діапазони звукових сигналів

Увесь звуковий діапазон можна умовно розділити на шість чітких частотних смуг, кожна з яких впливає на загальне звучання запису.

1. Низький бас, суббас (10 Гц ... 80 Гц) – це найнижчі звуки, від яких кімната резонує і починають гудіти дроти. Якщо звукова апаратура не відтворює дані частоти, слухач повинен відчувати втрату насиченості та глибини звуку. При звукозаписі та обробці звуку втрата цих частот викликає той самий ефект. В сигналі мови диктора ці частоти відсутні.

2. Верхній бас (80 Гц ... 200 Гц) – це верхні звуки басових інструментів і найнижчі ноти таких інструментів, як гітара. Якщо загубити цей регістр, то разом з ним втратиться відчуття сили сигналу. На даних частотах знаходиться основна енергія ритм-секції. Також на ці частоти приходиться основний тон чоловічого і деяких жіночих голосів.

3. Нижні середні, нижній діапазон середніх частот (200 Гц ... 2000 Гц) – в даній області знаходяться нижні гармоніки більшості музичних інструментів. Якщо дану область підняти надто сильно, то звук може набути "телефонний" характер. Якщо сильно підсилити смугу частот 500 ... 1000 Гц, то звучання нагадуватиме мідні духові інструменти, а якщо надто підняти частоти 1000 ... 2000 Гц, то звук стане "бляшаним". Надмірна гучність цієї смуги частот стомлює слух.

4. Верхні середні, верхня частина середньочастотного діапазону (2 кГц ... 4 кГц) – якщо надто сильно підняти цю смугу частот, то замаскуються звуки, що важливі для сприйняття мови. Мова диктора стане шепелявою, фонемі, які формуються губами, стануть невиразними. Зайве підсилення цієї смуги частот, особливо 3 кГц, помітно стомлює слух. У випадках, коли голос прихований під музикою, потрібно спробувати прибрати 3 кГц у музиці, а у голосі – додати. Так можна показати голос, не забираючи гучність музичної доріжки.

5. Діапазон присутності (4 кГц ... 6 кГц) – "відповідає" за прозорість та визначеність звучання мови. При підсиленні цього частого діапазону музика та мова здаються "ближчою" до слухача (рис.1.5). Зменшення рівня частоти 5 кГц робить звучання більш віддаленим.

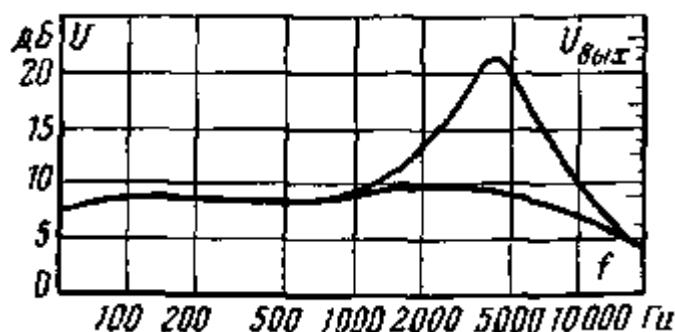


Рисунок 1.5 – Амплітудно-частотна характеристика фільтра присутності

6. Діамант, блиск (6 кГц ... 16 кГц) – задає блиск і прозорість звуку. Надто великий підйом в цій області частот може перебільшити шиплячі фонemi (сібілянти) у дикторів і співаків.

1.5.3 Амплітудно-частотні характеристики фільтрів

Найчастіше використовуються в еквайзерах фільтри з наступними амплітудно-частотними характеристиками АЧХ [14]:

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

На рис.1.6 наведені АЧХ Low pass та High pass фільтрів.

На рис.1.7 наведені АЧХ Peaking або bell фільтрів.

На рис.1.8 наведені АЧХ Low shelf та High shelf фільтрів.

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

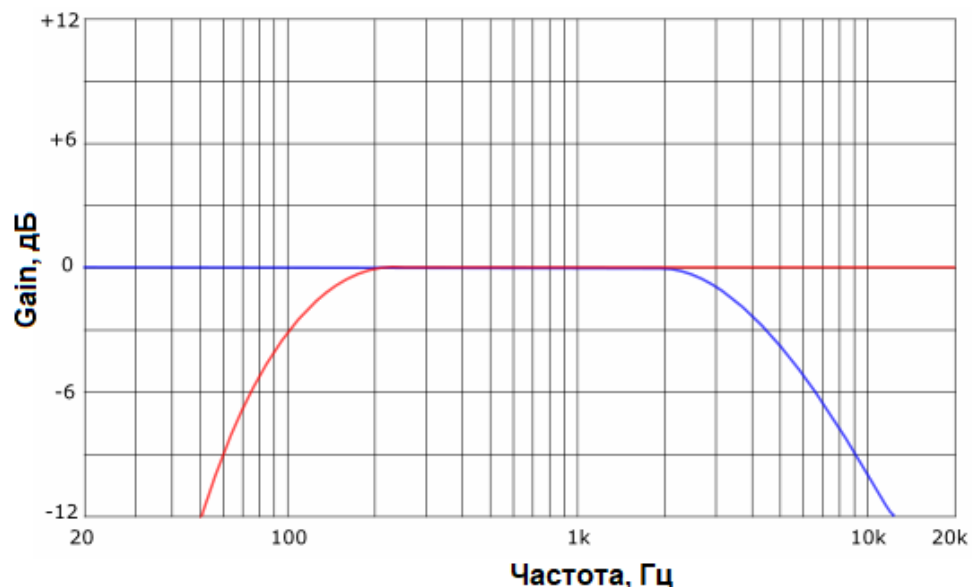


Рисунок 1.6 – АЧХ Low pass та High pass фільтрів

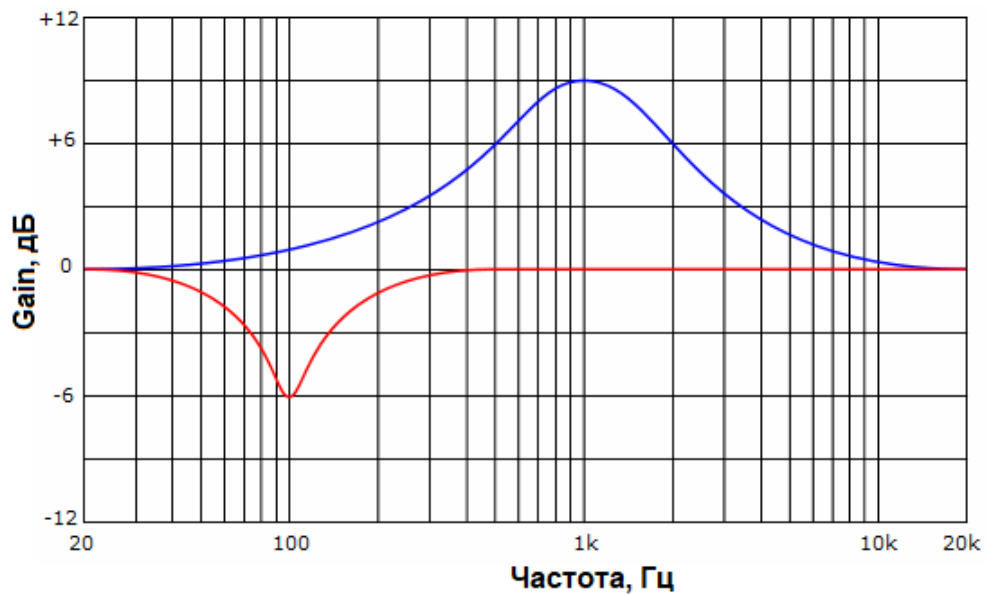


Рисунок 1.7 – АЧХ Peaking або bell фільтрів

Гнучкість цифрового еквайзера обмежено тільки програмним кодом, у цифрових еквайзерів є можливість налаштувати дуже велику крутизну спаду АЧХ, і в деяких пристроях створювати майже необмежену кількість смуг [14].

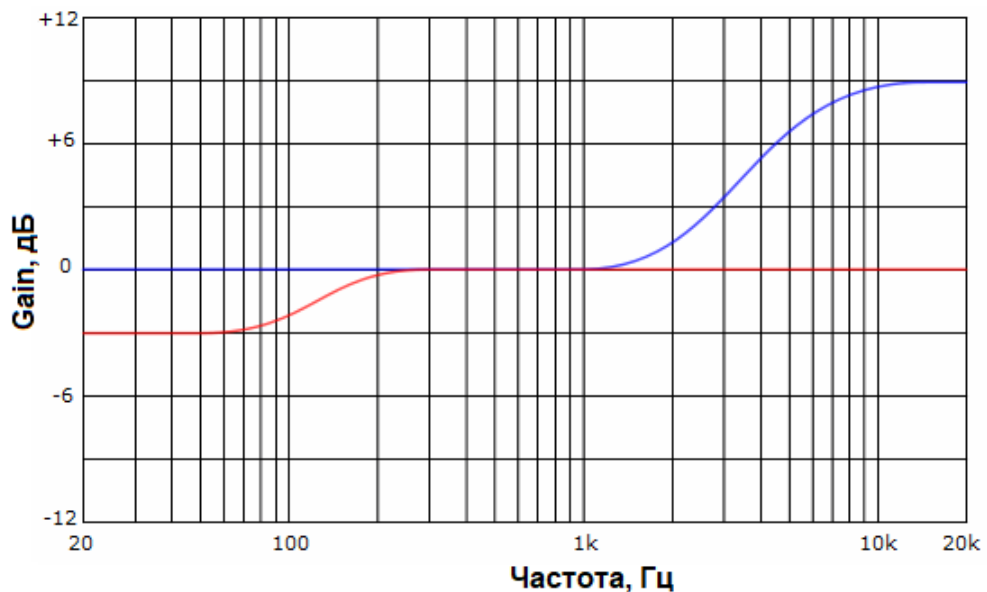


Рисунок 1.8 – АЧХ Low shelf та High shelf фільтрів

1.5.4 Фазо-частотні характеристики фільтрів

Аналогові еквайзери дають частотно-залежний зсув у часі при проходженні сигналів. Тобто, різні компоненти різних частот сигналу

зсуваються на різний час. Найчастіше, це небажаний ефект, оскільки якщо на вхід поступає імпульсний сигнал (атака), то і на виході потрібно отримати чіткий імпульс, що не розмазаний у часі [14].

Фазочастотна характеристика (фазова характеристика, ФЧХ, phase response) показує, як змінюється фаза сигналів при проходженні крізь еквалайзер. Якщо фаза міняється на значення, пропорційне частоті, це відповідає простому зміщенню сигналу в часі, без спотворення форми сигналу. Еквалайзери, які реалізують подібну ФЧХ, називаються linear-phase EQ (еквалайзер з лінійною фазою). Усі аналогові еквалайзери мають ФЧХ нелінійну, а цифрові еквалайзери можуть мати ФЧХ як лінійну, так і нелінійну.

В більшості аналогових еквалайзерів є можливість обримати ФЧХ за відомою АЧХ. При цьому найбільші спотворення ФЧХ відбуваються у місцях швидких змін АЧХ. Це означатиме, що чим сильніше втручання в частотну характеристику, тим сильніше проявлятимуться фазові спотворення. Звукоінженери часто говорять, що еквалайзер «крутить» фазу.

Для умови відсутності фазових спотворень треба, щоби груповий час затримки (тобто зміщення часу різних частотних компонентів сигналу) у смузі частот був незмінним [14]

$$t_{gp} = \frac{d\varphi(\omega)}{d\omega} = const, \quad (1.3)$$

де $\omega = 2\pi f$ – кругова частота.

Тобто ФЧХ повинна лінійно залежати від частоти. Фазові спотворення описуються різницею у часі пробігу по каналу сигналів середньої частоти і крайніх частот діапазону:

$$\Delta t_H = \left. \frac{d\varphi}{d\omega} \right|_{\omega_H} - \left. \frac{d\varphi}{d\omega} \right|_{\omega_{cp}}, \quad \Delta t_G = \left. \frac{d\varphi}{d\omega} \right|_{\omega_G} - \left. \frac{d\varphi}{d\omega} \right|_{\omega_{cp}}. \quad (1.4)$$

У відповідності рекомендації ITU-T (International Telecommunication Union – Telecommunication sector) середня частота обирається 800 Гц.

Цифрові еквалайзери можуть бути з лінійною ФЧХ і постійним груповим часом затримки (ГЧЗ), такі еквалайзери називають лінійними.

Недоліки лінійних еквалайзерів [14]:

- наявність пульсацій АЧХ в смузі пропускання;
- наявність неповного подавлення в смузі подавлення;
- обмеження крутизни зміни АЧХ;
- наявність затримки сигналу.

Затримка сигналу в лінійних еквалайзерів на практиці зустрічається не завжди, багато сучасних аудіоредакторів і секвенсорів компенсують дану затримку, це дозволяє застосовувати їх поряд з нелінійними еквалайзерами.

1.5.5 Імпульсний відгук фільтрів

Імпульсний відгук (імпульсна характеристика) фільтра – це вихідний сигнал фільтра при надходженні на його вхід одиничного імпульсу.

Спостерігаючи імпульсний відгук можна зрозуміти, як еквалайзер буде діяти на різкі сплески в звуковому сигналі. Якщо максимум імпульсного відгука оточений коливаннями з двох сторін, то можна очікувати, що і різкі сплески звукового сигналу будуть оточені коливаннями – даний ефект називається «дзвоном» (ringing) фільтру.

В лінійних еквалайзерах дзвін в однаковій мірі присутній до і після імпульсу. В більшості нелінійних еквалайзерів дзвін зосереджено після імпульсу, це часто є перевагою, оскільки за різкими атаками звуку зазвичай слідує порівняно плавне спадання, яке з великою вірогідністю психоакустично замаскує дзвін фільтра. Випереджуючий дзвін лінійного еквалайзера більш ймовірно буде помітний перед різкими ударами.

На рис.1.9 показані імпульсні характеристики нелінійного (а) і лінійного (б) фільтрів.

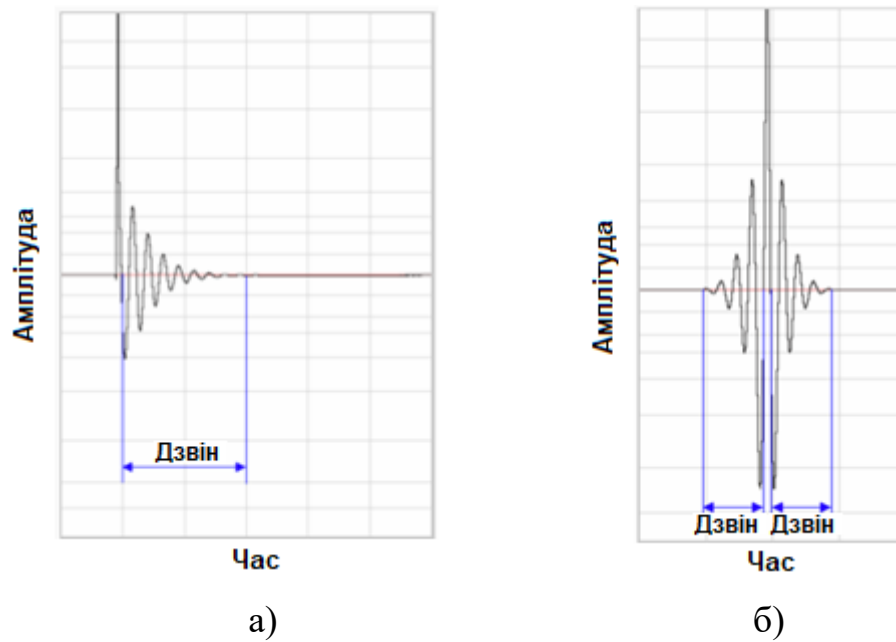


Рисунок 1.9 – Імпульсна характеристика нелінійного (а) і лінійного (б) фільтрів

Величина дзвону (його відносний рівень і тривалість у часі) залежить безпосередньо від АЧХ та ФЧХ фільтру. Чим крутіші зміни в АЧХ потрібні від еквайзера, тим більше дзвону буде на виході у часовій області. При цьому дзвін буде зосереджуватися в околицях частот «зламу» та найбільшої кривизни АЧХ та ФЧХ.

1.6 Висновки по розділу 1

Об'єкт дослідження – процеси спотворення і корекції частотних характеристик звукозаписів мови диктора.

Мета роботи – розробити методику оцінки нерівномірності АЧХ в тракці звукозапису без використання спеціальних випробувальних сигналів з подальшим виправленням цих нерівномірностей коригуючим еквайзером.

Задачі дослідження.

1. Виконати теоретичну оцінку факторів, що впливають на наскрізну АЧХ тракту формування, обробки, передачі і відтворення мовного сигналу.

2. Визначити головні фактори, що впливають на якісні показники (натуральність) звучання мови диктора.

3. Отримати вираз для оціночної кривої АЧХ еквайзера корекції.

4. Отримати теоретичні оцінки характерних особливостей спектрів мови диктора і швидкості спадання спектральної щільності потужності.

5. Провести уточнення теоретичних результатів шляхом експериментального дослідження аудіо записів реальних ефірів на радіо, телебаченні і в Ютубі.

6. Провести експериментальне дослідження корекції АЧХ мови диктора для оцінки можливостей і обмежень автоматизованої системи корекції.

Дана кваліфікаційна робота виконана на кафедрі МІРЕС ХНУРЕ. На кафедрі проводяться дослідження в таких наукових областях, як виявлення та розпізнавання БПЛА за результатами акустичного спостереження [15-18], створення систем зондування атмосфери за допомогою акустичних хвиль [19-22]. Цілий ряд студентських доповідей [23-26] і атестаційних робіт магістрів минулих років [27-31] присвячені дослідженню систем озвучення приміщень. Отже, дослідження в даній роботі відповідають традиційному напрямку наукових робіт колективу кафедри МІРЕС.

2 ТЕОРЕТИЧНЕ ОБГРУНТУВАННЯ ПРИНЦИПІВ ЕКВАЛІЗАЦІЇ МОВНИХ СИГНАЛІВ

2.1 Тракт формування, обробки, передачі і відтворення мовного сигналу

Тракт передачі мови (ТПМ), узагальнена схема якого наведена рис. 2.1, складається з приміщення студії (Пр), мікрофона М, звукозаписувальної апаратури (ЗвЗ), каналу зв'язку (КЗ), звуковідтворювальної апаратури (ЗвВ) та гучномовця (Г).

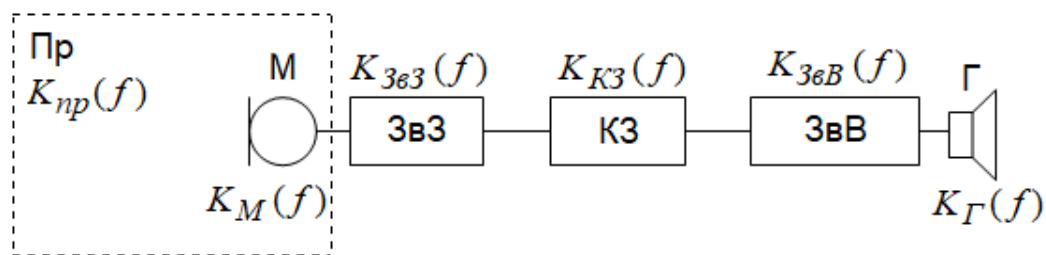


Рисунок 2.1 – Структурна схема тракту передачі мови

Основними критеріями якості тракту передачі мови є [32]:

- розбірливість;
- гучність;
- натуральність.

При цьому гучність не є самодостатнім параметром – вона використовується спільно з розбірливістю і визначає бажаний (комфортний) рівень сигналів, що приймаються. Навіть із особистої практики кожної людини відомо, що надто низький рівень гучності призводить до зниження розбірливості. Дослідження показують, що розбірливість знижується і за надто високого рівня гучності мови.

Натуральність звучання – це здатність системи відтворювати як сенс промови, а також і її тембр, індивідуальні особливості промови диктора. Для трактів зв'язку (телефонний зв'язок, радіозв'язок) ця характеристика є другорядною. Але в трактах звукового, радіо- і телевізійного мовлення, коли стоїть завдання високоякісного відтворення промови диктора (або співу), ця

характеристика є дуже важливою. У нашому випадку натуральність мови також є першорядною, тому що стоїть завдання передачі особливих і характерних особливостей мови диктора, точної передачі тембрального зафарблення.

Таким чином, натуральність мови є основним параметром, що характеризує тракт передачі в нашому випадку.

Якщо вважати тракт передачі лінійним, то загальна (наскрізна) АЧХ тракту виражається як добуток АЧХ окремих ланок тракту;

$$K(f) = \frac{P_{вих}(f)}{P_{вх}(f)} = K_{np}(f) \cdot K_M(f) \cdot K_{звз}(f) \cdot K_{КЗ}(f) \cdot K_{звВ}(f) \cdot K_{Г}(f), \quad (2.1)$$

де $P_{вих}$, $P_{вх}$ – звуковий тиск на вході і виході системи відповідно;

$K_{np}(f)$ – АЧХ звучання приміщення;

$K_M(f)$ – АЧХ мікрофона;

$K_{звз}(f)$, $K_{звВ}(f)$ – АЧХ звукозаписуючого і звуковідтворюючого тракту відповідно;

$K_{КЗ}(f)$ – АЧХ каналу зв'язку;

$K_{Г}(f)$ – АЧХ гучномовця.

В цифрових системах передавання інформації можна вважати, що

$$K_{КЗ}(f) = 1 \quad (2.2)$$

у смузі частот звукового тракту 20 Гц ... 20 кГц. Задача користувача, який хоче забезпечити якісне відтворення звуку, щоби виконувалася умова

$$K_{звВ}(f) \cdot K_{Г}(f) \approx 1. \quad (2.3)$$

Умову (2.3) користувач може виконати за рахунок використання високоякісних гучномовців і еквайзера в звуковідтворюючому тракті, який

компенсує нерівномірність АЧХ гучномовця і приміщення, де прослуховується звукозапис.

Отже, з урахуванням (2.2), (2.3) вираз (2.1) можемо записати у вигляді

$$K(f) = K_{np}(f) \cdot K_M(f) \cdot K_{ЗвЗ}(f). \quad (2.4)$$

При умові рівномірності наскрізної АЧХ $K(f)=1$ у смузі частот звукового тракту 20 Гц ... 20 кГц. Тобто

$$K_{np}(f) \cdot K_M(f) \cdot K_{ЗвЗ}(f) = 1. \quad (2.5)$$

На практиці, навіть в умовах студії звукозапису,

$$K_{np}(f) \neq 1, \quad K_M(f) \neq 1, \quad (2.6)$$

тому виконати умову (2.5) можна лише за рахунок еквалізації тракту звукозапису (ЗвЗ). При цьому АЦХ еквалайзера повинна відповідати умові:

$$K_{ЗвЗ}(f) = \frac{1}{K_{np}(f) \cdot K_M(f)}. \quad (2.7)$$

Не зважаючи на уявну простоту виразу (2.5) на практиці налаштування еквалайзеру виглядає набагато складніше. Справа в тому, що заздалегідь невідомий вигляд функцій $K_{np}(f)$ та $K_M(f)$. Навіть якщо застосовується мікрофон з відомою АЧХ, реальний вигляд функції $K_M(f)$ відрізнятиметься від паспортних даних за рахунок випадкового розміщення диктора відносно мікрофона, як за відстанню, так і за кутом. Практика використання мікрофонів підтверджує цю залежність.

АЧХ звучання приміщення теж залежить від місць розташування мікрофону і джерела звуку (диктора). Тому стоїть некоректна в математичному сенсі задача визначення функції АЧХ еквалайзера в умовах апіорної невизначеності функцій $K_{np}(f)$ та $K_M(f)$.

Єдиним «випробувальний» сигналом, що є в системі, це мова диктора. Спектр сигналу на виході системи:

$$S(f) \approx K_{np}(f) \cdot K_M(f) \cdot S_\partial(f), \quad (2.8)$$

де $S_\partial(f)$ – спектр сигналу диктора.

Насправді, точний вигляд функції $S_\partial(f)$ теж є апіорі невідомим. Ми можемо користуватися лише оціночним значенням функції $S_\partial(f)$ за тривалий час, яку позначимо як $\hat{S}_\partial(f)$ і будемо використовувати як основу для налаштування еквалайзера.

Поєднуючи вирази (2.7) і (2.8), запишемо

$$K_{звз}(f) = \frac{\hat{S}_\partial(f)}{S(f)}, \quad (2.9)$$

де $\hat{S}_\partial(f)$ – оціночний (або цільовий) спектр мови диктора, яким він повинен бути в ідеалі;

$S(f)$ – реальний спектр звукозапису мови диктора, який він є.

Формула (2.9) дає оціночний вигляд АЧХ коригуючого еквалайзера і не виключає подальшої тонкої ручної корекції.

Зрозуміло, що вигляд функцій $\hat{S}_\partial(f)$ і $S(f)$ залежатиме від інтервалу спостереження. Тому одним із питань є визначення інтервалу інтегрування при визначенні спектрів $\hat{S}_\partial(f)$ і $S(f)$.

2.2 Тонка структура спектра і середня спектральна щільність потужності

Розглянемо спектральні властивості звуків мови. Спектри голосних звуків є (у першому наближенні) періодичній послідовністю спектральних піків. Період проходження цих піків називають частотою основного тону. Виражені сплески рівня «обвідної» спектральних піків називають «формантами» (рис. 2.2).

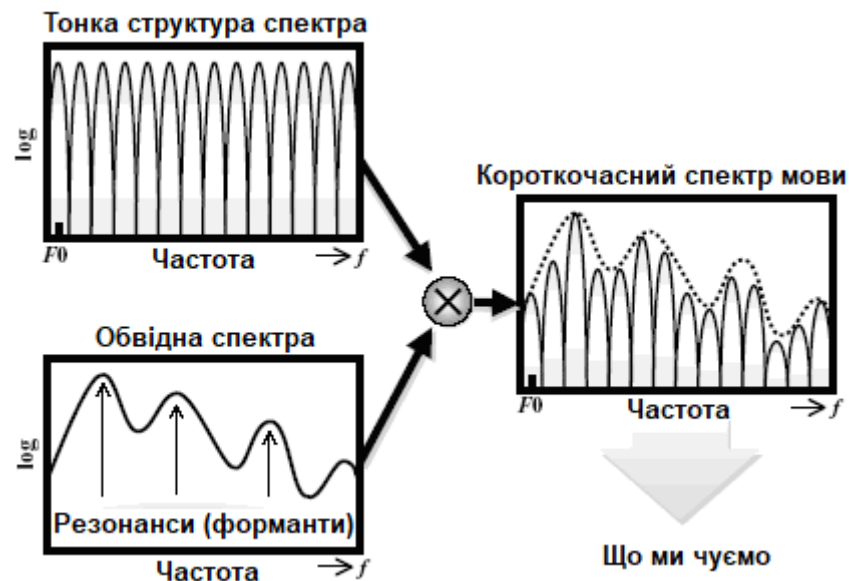


Рисунок 2.2 – Вигляд спектру голосного звуку

Корисна інформація про голосний звук мови міститься в описі відповідних формант. Прийнято кожному форманту описувати її граничними частотами. У російській досить обмежитися однією-двома формантами, щоб досягти прийнятної розбірливості промови.

Спектри приголосних звуків чи цілком суцільні, тобто. Не містять дискретних компонентів, чи суцільні у окремих смугах частот. Ці спектри містять локальні сплески. Деякі є формантами, деякі – ні.

Щоб вирішити, які сплески рівня спектра є формантами, слід пам'ятати, що фізична природа формант – явище резонансу порожнинах глотки і носоглотки (рис. 2.3).

В окремих звуках можна побачити до 6 спектральних підйомів. До формантів належать лише ті, які зумовлені явищем резонансу в мовному

апараті людини. Частина формант (як вказувалося – одна-дві українською мовою) забезпечують розбірливість промови, інша частина забезпечує індивідуальність голосу диктора, що може бути використане завдання розпізнавання голосу (ідентифікації) диктора [32].

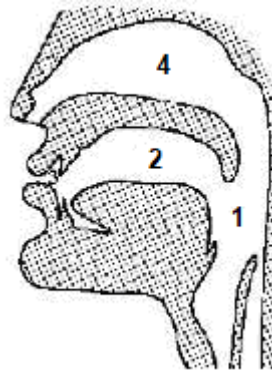


Рисунок 2.3 – Порожнини глотки (1) та носоглотки (2-4)

Форманти звуків мови розташовані в області частот від 200 до 8600 Гц. Однак переважна частина формант звуків мови знаходиться в межах від 300 до 3000 Гц, тому таку смугу зазвичай вважають достатньою для хорошої розбірливості мови. Спектральні відмінності між звуками мови є головними, хоч і не єдиними. Наприклад, при розпізнаванні приголосних звуків важливі і часові характеристики.

Інтегральні спектральні властивості мови. Як випливає з назви, інтегральні спектральні властивості мови характеризують характеристики мовного процесу загалом. Сюди відносять [32]:

- спектр мови $S_M(f)$;
- спектр формант $S'_M(f)$;
- відносну зустрічальність формант по спектру $\Delta A\%(f)$.

Спектром мови називають оцінку спектральної щільності потужності мовного сигналу $S_M(f)$, обчислену за відрізком мовного сигналу значної протяжності (більше за хвилину). Спектр мови характеризує розподіл потужності мовного сигналу частотою.

Спектром формант $S'_m(f)$ називають залежність найбільш ймовірного рівня формант від частоти. Щоб оцінити спектр формант, необхідно також мати відрізок мови великої протяжності. Спектр формант на всіх частотах менший за спектр мови (рис. 2.4) [32]:

$$S_m(f) < S'_m(f). \quad (2.10)$$

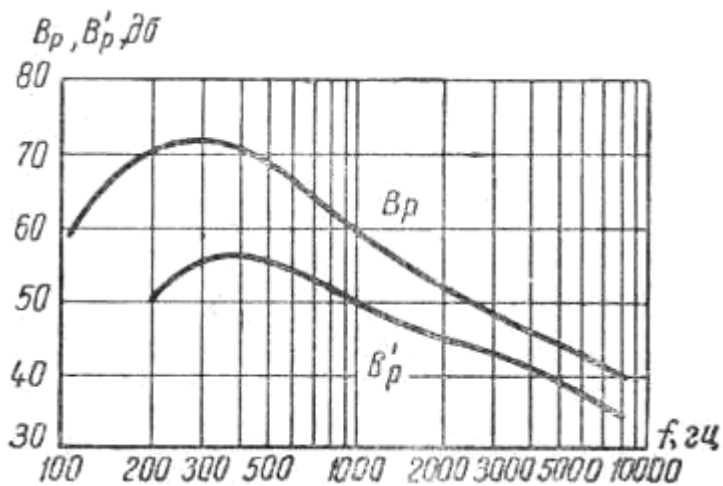


Рисунок 2.4 – Співвідношення спектра мови $S_m(f)$ та спектра формант $S'_m(f)$

Відносна зустрічальність формант за спектром можна оцінити так. Розіб'ємо весь діапазон частот на смужки, наприклад, по 100 Гц, і підрахуємо відносне число формант (%) у кожній смугі. Результат такого підрахунку дасть нам криву (рис. 2.5) [32].

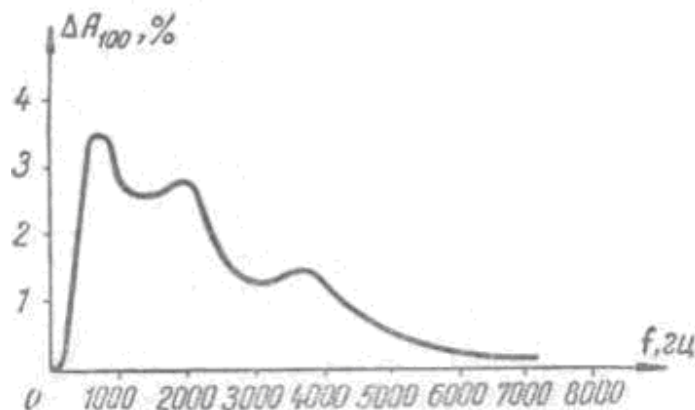


Рисунок 2.5 – Відносна зустрічальність формант $\Delta A\%(f)$

Проведемо теоретичні оцінки середньої швидкості спадання спектру мовних сигналів, наведених на рис. 2.4. Для цього апроксимуємо пласку частину обвідної спектру та обвідної формант лінійною функцією, враховуючи, що графіки побудовані в логарифмічному масштабі (рис.2.6).

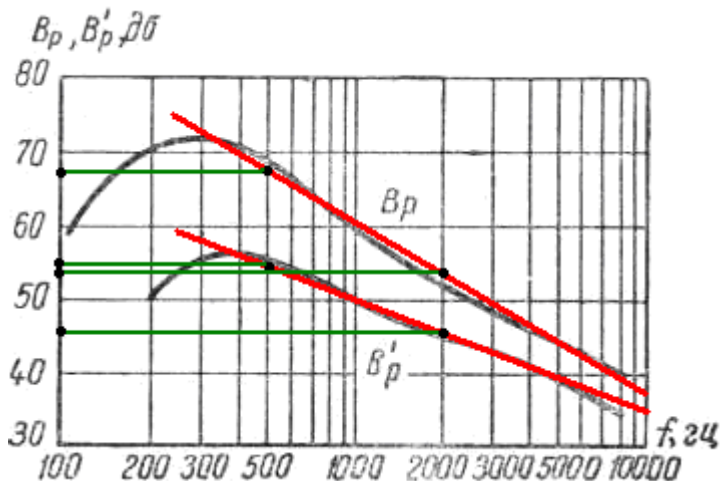


Рисунок 2.6 – Апроксимація спектру мови $S_M(f)$ та спектру формант $S'_M(f)$ лінійними функціями

Розрахуємо коефіцієнти нахилу для лінійних оцінок спектру мови $S_M(f)$ та спектру формант $S'_M(f)$:

$$k_M = \frac{S_M(f_2) - S_M(f_1)}{\log_2(f_2 / f_1)}, \quad (2.11)$$

$$k'_M = \frac{S'_M(f_2) - S'_M(f_1)}{\log_2(f_2 / f_1)}, \quad (2.12)$$

де S_M, S'_M – рівні спектральної щільності, вимірювані в дБ,

$\log_2(f_2 / f_1)$ – частотна відстань в октавах.

Випишемо в табл.2.1 відповідні дані з графіків на рис.2.5. Підставимо в вирази (2.11), (2.12) виписані дані з табл.2.1. Тоді отримаємо такі значення лінійної швидкості спадання спектральної щільності потужності обвідної мовних сигналів $S_M(f)$ і обвідної спектру формант $S'_M(f)$:

Таблиця 2.1 – Дані для розрахунку спектральних нахилів

f , Гц	500	2000
$S_M(f)$, дБ	78	55
$S'_M(f)$, дБ	53	46

$$k_M = \frac{78 - 55}{\log_2(2000/500)} = \frac{23}{2} = 11,5 \text{ дБ/окт.}, \quad (2.14)$$

$$k'_M = \frac{53 - 46}{\log_2(2000/500)} = \frac{7}{2} = 3,5 \text{ дБ/окт.} \quad (2.15)$$

Виводячи середнє значення між швидкостями спадіння спектру мови (2.14) та формант мови (2.15) можна отримати значення

$$k = \frac{k_M + k'_M}{2} = \frac{11,5 + 3,5}{2} = 7,5 \text{ дБ/окт.} \quad (2.16)$$

Отже, в результаті теоретичного аналізу ми отримали оціночне значення швидкості спаду спектру мовного сигналу з частотою. Слід очікувати, що при такій швидкості спаду мова буде мати натуральне звучання. Дана швидкість спаду має бути для частот від 200...300 Гц і до верхніх частот звукового діапазону. Область нижніх частот, нижче 200...300 Гц доцільно не чіпати і не оброблювати еквалайзером.

Для уточнення отриманих результатів доцільно провести експериментальне дослідження аудіо записів реальних ефірів на радіо, телебаченні і в Ютубі. Це дасть можливість отримати рекомендації для приведення аудіо записів до поширених нині звичних для слухачів стандартів АЧХ.

Також важливо врахувати умови прослуховування. Реальне прослуховування виконується у умовах шуму. Тому, щоби звук залишався розбірливим, треба врахувати властивості слуху.

2.3 Врахування властивостей слуху

Для оцінювання розбірливості мови найбільше значення мають такі показники слухової системи людини, іменовані «постійними слухами» [32]:

- поріг чутності $\beta_0(f)$;
- логарифмічна ширина критичної смуги слуху $K(f)$;
- маскування слуху $M(f)$.

Поріг чутності

$$\beta_0(f) = 20 \lg \frac{P_{\min}(f)}{P_0}, \quad (2.17)$$

це мінімальний звуковий тиск, нижче якого вухо не сприймає звук (рис. 2.7) [32]. Виражається в децибелах, по відношенню до тиску $P_0 = 2 \cdot 10^{-5}$ Па, що відповідає граничній величині тиску звуку на частоті 1000 Гц.

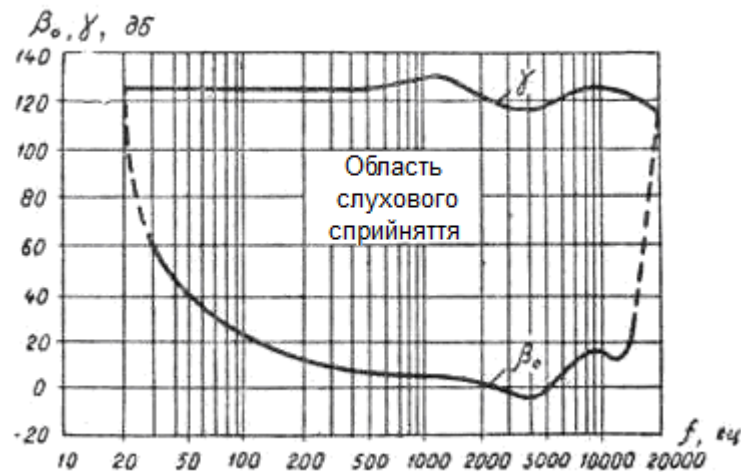


Рисунок 2.7 – Поріг чутності $\beta_0(f)$ та больовий поріг $\gamma(f)$

Ширина критичної смуги слуху – це здатність слухового апарату людини, який можна уподібнити гребінці фільтрів. Наприклад, на частоті

100 Гц критична смуга слуху близько 100 Гц, але в частоті 8000 Гц – близько 600 Гц (рис. 2.8) [32].

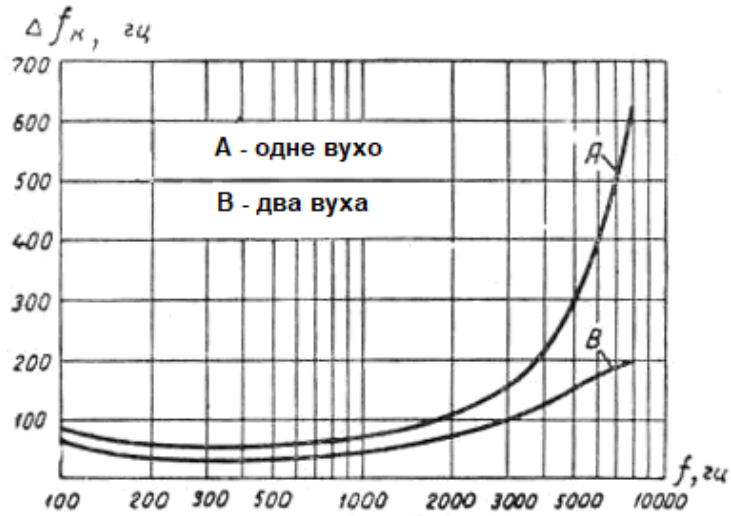


Рисунок 2.8 – Критична смуга слуху $\Delta f_k(f)$

Для зручності розрахунків вводять поняття логарифмічної критичної смуги слуху (рис.2.8) [32]:

$$k(f) = 10 \lg \Delta f_k(f). \quad (2.18)$$

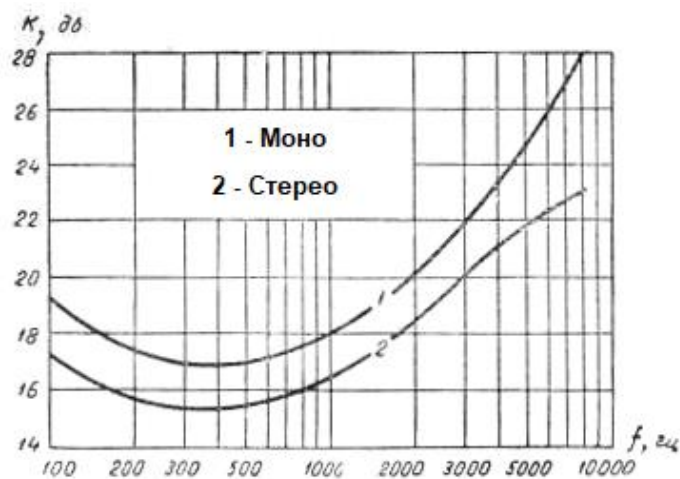


Рисунок 2.8 – Логарифмічна критична смуга слуху

Маскування слуху – це явище ослаблення чутності або повного пропадання корисного звуку на тлі звуку, що заважає. Кількісно виражається як різниця:

$$M(f) = \beta_M(f) - \beta_0(f), \quad (2.19)$$

де $\beta_M(f)$ – поріг чутності за наявності завад звуку. На рис. 2.9 наведено сімейство індивідуальних кривих маскування для різних рівнів маскуючого сигналу N . Тут $\Delta z = z(i) - z(j)$ – різниця висот тону маскуючої $z(i)$ і маскуємої $z(j)$ компонент, причому висота тону z вимірюється в Барках [32]:

$$z = 13 \cdot \arctg(0.0076 \cdot f) + 3.5 \cdot \arctg[(f / 7500)]^2. \quad (2.20)$$

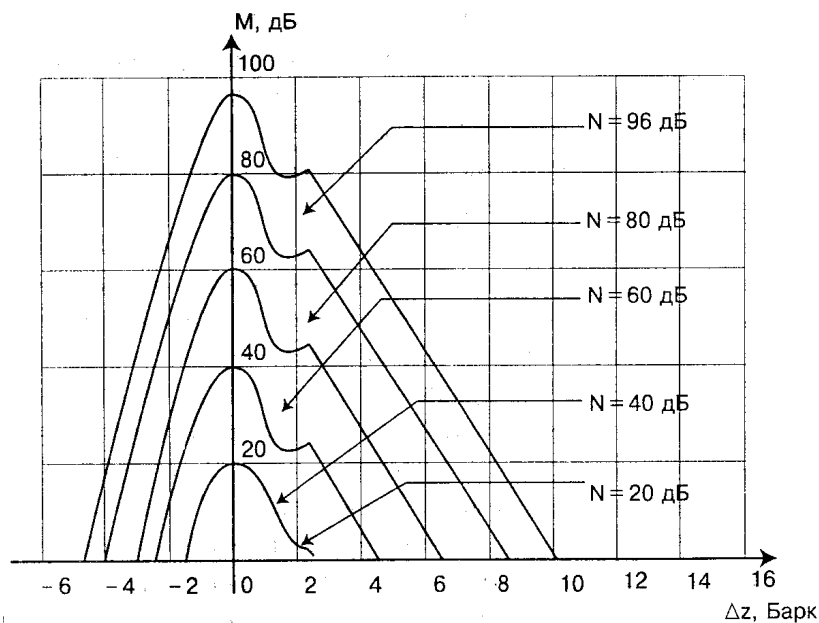


Рисунок 2.9 – Сімейство індивідуальних кривих маскування

При розрахунку розбірливості доводиться мати справу із частотно-залежними функціями. Тому результати кількісного розрахунку різних ділянок спектра різні. Зважаючи на це завдання розрахунку вирішується поділом діапазону частот, що використовується для передачі мови, на вузькі смуги, всередині яких можна не зважати на зазначену частотну залежність і відносити отримані результати до середньої частоти смуги. Далі обчислюється перебірливість кожної смуги частот, а загальна розбірливість перебуває підсумовуванням «смугових» розбірливостей.

2.4 Висновки за розділом 2

В результаті теоретичного обґрунтування принципів еквалізації мовних сигналів показано, що основним показником якості передачі у вирішуваній задачі є натуральність звучання, що забезпечується ідентичністю тембру джерела і відтворюваного звука.

Розглянуто структурну схему наскрізного тракту запису і відтворення звуку. Отримано вираз для АЧХ еквалайзера корекції. Не зважаючи на уявну простоту виразу (2.5), проблема полягає в тому, що заздалегідь невідомий вигляд АЧХ приміщення і мікрофону. Навіть якщо застосовується мікрофон з відомою АЧХ, реальний вигляд АЧХ відрізнятиметься від паспортних даних за рахунок невідомого розміщення диктора відносно мікрофона. АЧХ звучання приміщення теж залежить від місць розташування мікрофону і джерела звуку (диктора). Тому треба вирішити некоректну в математичному сенсі задачу визначення АЧХ еквалайзера в умовах апріорної невизначеності функцій $K_{np}(f)$ та $K_M(f)$.

Єдиним «випробувальний» сигналом, що є в системі, це мова диктора. Ми можемо користуватися лише оціночним значенням спектра голосу диктора, яку і будемо використовувати як основу для налаштування еквалайзера.

Проведено теоретичні оцінки середньої швидкості спадання спектру мовних сигналів. Для цього було апроксимовано пласку частину обвідної спектру та обвідної формант лінійною функцією, враховуючи, що графіки побудовані в логарифмічному масштабі. Виводячи середнє значення між швидкостями спадання спектру мови та формант мови отримано значення 7,5 дБ/окт.

Слід очікувати, що при такій швидкості спаду мова буде мати натуральне звучання. Дана швидкість спаду має бути для частот від 200...300 Гц і до верхніх частот звукового діапазону. Область нижніх частот, нижче 200...300 Гц доцільно не чіпати і не оброблювати еквалайзером.

Для уточнення отриманих результатів доцільно провести експериментальне дослідження аудіо записів реальних ефірів на радіо, телебаченні і в Ютубі. Це дасть можливість отримати рекомендації для приведення аудіо записів до поширених нині звичних для слухачів стандартів АЧХ.

Також важливо врахувати умови прослуховування. Реальне прослуховування виконується у умовах шуму. Тому, щоби звук залишався розбірливим, треба врахувати властивості слуху.

3 ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ СПЕКТРІВ МОВНИХ СИГНАЛІВ І ЇХ АВТОМАТИЧНОЇ ЕКВАЛІЗАЦІЇ

3.1 Обґрунтування методу вимірювання спектру мовних сигналів

Спектроаналізатор – прилад для вимірювання та відображення спектра сигналу – розподіл енергії сигналу по частотах. У цьому розділі розглядаються основні види аналізаторів спектру та ілюструється їх застосування для редагування та реставрації звуку. Особлива увага приділяється сучасним аналізаторам, що ґрунтуються на FFT – швидкому перетворенні Фур'є.

Традиційно в цифровому звукозаписі аудіодоріжка представляється у вигляді осцилограми, що відображає форму звукової хвилі (waveform), тобто залежність амплітуди звуку від часу. Таке представлення є досить наочним для звукорежисерів і звукоінженерів: осцилограма дозволяє побачити основні події в звуку, такі як зміни гучності, паузи між частинами твору і навіть окремі ноти в сольному записі інструменту. Але одночасне звучання кількох інструментів на осцилограмі "змішується" і візуальний аналіз сигналу стає складним.

Перші звукові аналізатори спектру поділяли сигнал на частотні смуги за допомогою набору аналогових фільтрів. Дисплей такого аналізатора (рис. 3.1) показує рівень сигналу у множині частотних смуг, що відповідають фільтрам.

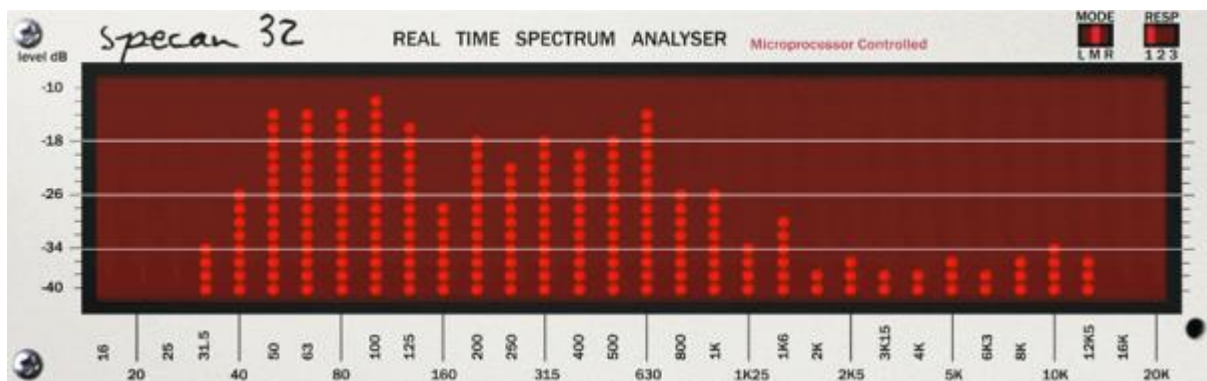


Рисунок 3.1 – Третьоктавний аналізатор Spesca32, що емулює відомий апаратний прилад KlarkTeknik DN60

На рис. 3.2 наведено приклад частотних характеристик смугових фільтрів в аналізаторі, що відповідає радянському стандарту ГОСТ 17168-82. Такий аналізатор називається третьоктавним, так як у кожній октаві частотного діапазону є три смуги. Видно, що частотні характеристики смугових фільтрів перекриваються; їхня крутість залежить від порядку використовуваних фільтрів.

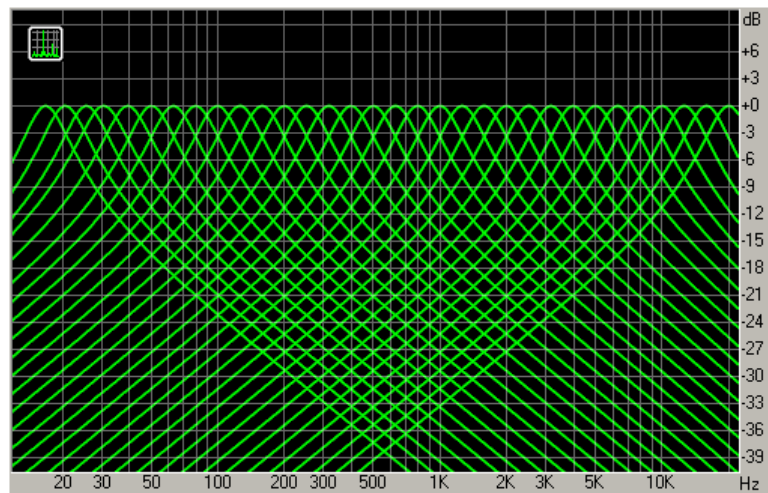


Рисунок 3.2 – Частотні характеристики фільтрів третьоктавного спектроаналізатора

Важливою властивістю спектроаналізатора є балістика – інерційність вимірювачів рівня частотних смуг. Вона може регулюватися задачею швидкості наростання (атаки) та спаду рівня. Типовий час атаки та спаду в такому аналізаторі – близько 200 та 1500 мс.

Смужні спектроаналізатори часто використовуються для налаштування АЧХ (амплітудно-частотної характеристики) акустичних систем на концертних майданчиках. Якщо на вхід такому аналізатору подати рожевий шум (має однакову потужність у кожній октаві), то дисплей покаже горизонтальну лінію, з можливою поправкою на варіацію шуму в часі. Якщо рожевий шум, проходячи через звукопідсилювальну систему залу, спотворився, то зміни його спектра буде видно на аналізаторі. При цьому аналізатор, як і наше вухо, буде малочутливим до вузьких провалів АЧХ (менше 1/3 октави).

Для побудови спектра в цифровому вигляді можна використовувати перетворення Фур'є.

Перетворення Фур'є – це математичний апарат розкладання сигналів на синусоїдальні коливання. Наприклад, якщо сигнал $s(t)$ безперервний і нескінченний за часом, його можна у вигляді інтеграла Фур'є:

$$s(t) = \frac{1}{2\pi} \int_0^{\infty} S_f \cos(2\pi f t + \varphi_f) df. \quad (3.1)$$

Інтеграл Фур'є збирає сигнал $s(t)$ з нескінченної множини синусоїдальних складових різних частот f , що мають амплітуди S_f і фази φ_f .

Насправді нас більше цікавить аналіз кінцевих за часом звуків. Оскільки мова не є статичним сигналом, її спектр змінюється у часі. Тому при спектральному аналізі нас цікавлять окремі короткі фрагменти сигналу. Для аналізу таких фрагментів цифрового аудіосигналу існує дискретне перетворення Фур'є:

$$s(n) = \sum_{k=0}^{N/2} S_k \cos \frac{2\pi k(n + \varphi_k)}{N}, \quad (3.2)$$

де N відліків дискретного сигналу $s(n)$ на інтервалі часу від 0 до $N-1$ синтезуються як сума кінцевого числа синусоїдальних коливань з амплітудами S_k та фазами φ_k .

Частоти цих синусоїд дорівнюють kF_s/N , де F_s – частота дискретизації сигналу, а N – число відліків вихідного сигналу $s(n)$ на аналізованому інтервалі. Набір коефіцієнтів S_k називається амплітудним спектром сигналу. Як видно з формули (3.2), частоти синусоїд, на які розкладається сигнал, рівномірно розподілені від 0 (постійна складова) до $F_s/2$ максимально

можливої частоти в цифровому сигналі. Таке лінійне розташування частот відрізняється від розподілу смуг третьоктавного аналізатора.

FFT (Fast Fourier Transform) – алгоритм швидкого обчислення дискретного перетворення Фур'є. Завдяки йому стало можливим аналізувати спектр звукових сигналів у реальному часі.

Розглянемо роботу типового FFT-аналізатора. На вхід надходить цифровий аудіосигнал. Аналізатор вибирає з сигналу послідовні інтервали («вікна»), на яких обчислюватиметься спектр, і вважає FFT у кожному вікні для отримання амплітудного спектру S_k . Обчислений спектр відображається у вигляді графіка залежності амплітуди від частоти (рис. 3.3). Аналогічно смуговим аналізаторам зазвичай використовується логарифмічний масштаб по осях частот і амплітуд. Але через лінійне розташування смуг FFT по частоті спектр може виглядати недостатньо детальним на нижніх частотах або надмірно осцилюючим на верхніх частотах.

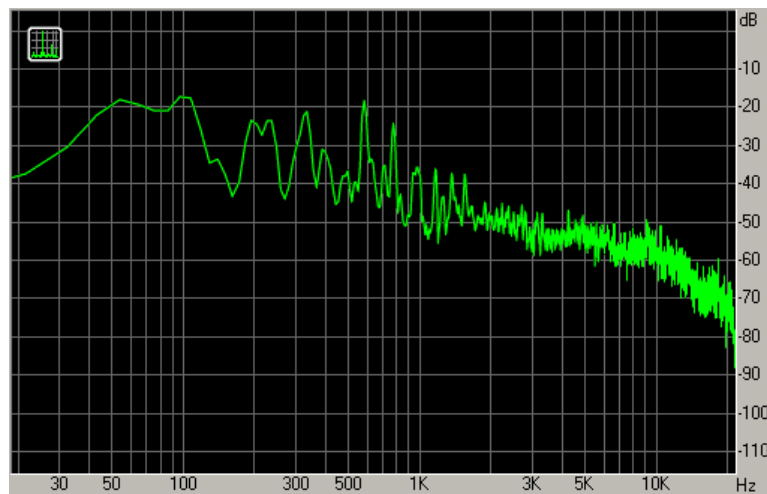


Рисунок 3.3 – Дисплей FFT-аналізатора

Якщо розглядати FFT як набір фільтрів, то, на відміну від смугових фільтрів третьоктавного аналізатора, FFT фільтри будуть мати однакову ширину в герцах, а не в октавах. Тому рожевий шум на FFT-аналізаторі буде вже не горизонтальною лінією, а похилою зі спадом 3 дБ/окт. Горизонтальною лінією на FFT-аналізаторі буде білий шум – він містить рівну енергію у рівних лінійних частотних інтервалах.

Параметр N – число аналізованих відліків сигналу – має вирішальне значення виду спектра. Чим більше N , тим щільніше сітка частот, якими FFT розкладає сигнал, і більше деталей по частоті видно у спектрі. Для досягнення вищої частотної роздільної здатності доводиться аналізувати довші ділянки сигналу. Якщо сигнал у межах вікна FFT змінює свої властивості, спектр буде відображати деяку усереднену інформацію про сигнал з усього інтервалу вікна.

Коли потрібно проаналізувати швидкі зміни сигналу, довжину вікна N вибирають маленькою. І тут розрізнення аналізу за часом збільшується, а, по частоті – зменшується. Таким чином, розрізнення аналізу за частотою обернено пропорційне розрізненню за часом. Цей факт називається співвідношенням невизначеності.

Для отримання більш точних оцінок спектру застосовують вагові вікна. Один із найпростіших звукових сигналів – синусоїдальний тон. Його спектр на FFT-аналізаторі залежить від частоти тону. Відомо, що FFT розкладає сигнал не за тими частотами, які насправді є у сигналі, а за фіксованою рівномірною сіткою частот. Наприклад, якщо частота дискретизації дорівнює 48 кГц і розмір вікна FFT вибрано 4096 відліків, то FFT розкладає сигнал по 2049 частот: 0 Гц, 11.72 Гц, 23.44 Гц, ..., 24000 Гц.

Якщо частота тону збігається з однією із частот сітки FFT, то спектр буде виглядати "ідеально": єдиний гострий пік вкаже на частоту та амплітуду тону (рис. 3.4, білий графік).

Якщо ж частота тону не збігається з жодною із частот сітки FFT, то FFT "збере" тон із наявних у сітці частот, скомбінованих з різними вагами. Графік спектра у своїй розмивається по частоті (рис. 3.4, зелений графік). Таке розмиття зазвичай небажане, оскільки воно може закрити собою слабкіші звуки на сусідніх частотах. Можна також помітити, що амплітуда максимуму зеленого графіка нижче за реальну амплітуду аналізованого тону. Це з тим, що потужність аналізованого тону дорівнює сумі потужностей коефіцієнтів спектра, у тому числі цей тон складено.

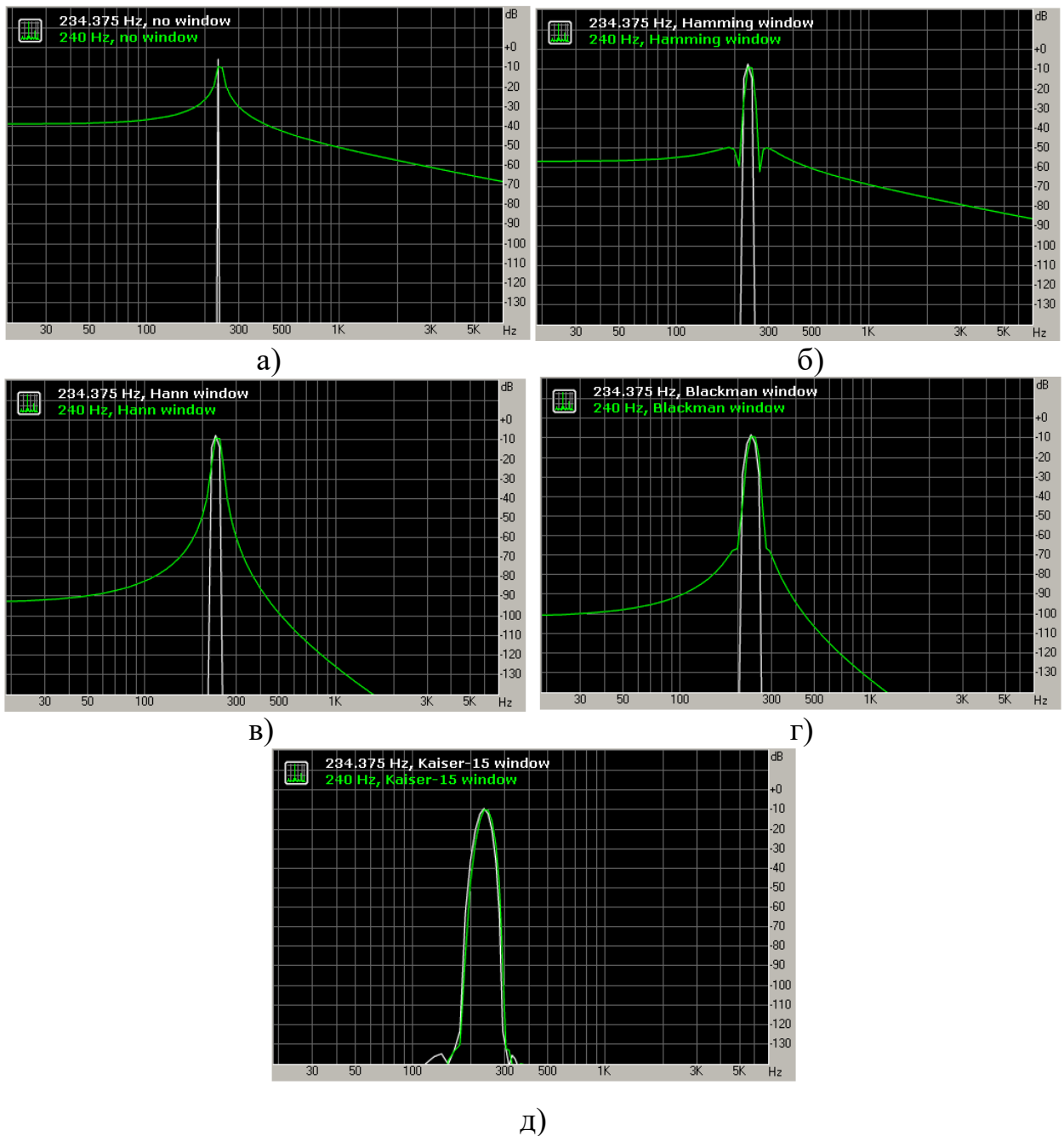


Рисунок 3.4 – Спектр синусоїдального тону різних частот з ваговими вікнами (зелене) та без них (біле), а – прямокутне вікно, б – вікно Хеммінга, в – вікно Хана, г – вікно Блекмана, д – Вікно Кайзера

Часто виникає необхідність простежити, як спектр сигналу змінюється у часі. FFT-аналізатори допомагають зробити це у реальному часі під час відтворення сигналу. Однак у ряді випадків виявляється зручна візуалізація зміни спектра у всьому звуковому уривку відразу. Таке уявлення сигналу називається спектрограмою. Для її побудови застосовується віконне

перетворення Фур'є: спектр обчислюється від послідовних вікон сигналу (рис. 3.5), і кожен із спектрів утворює стовпець в спектрограмі.

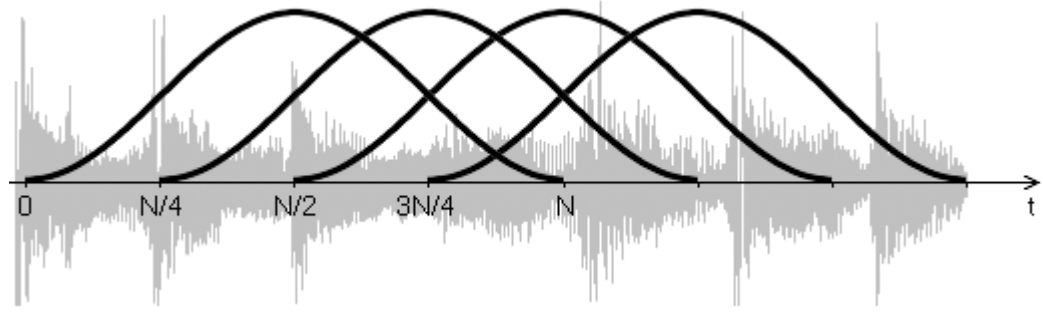


Рисунок 3.5 – Обчислення спектрограми сигналу

По горизонтальній осі спектрограми відкладається час, вертикальною – частота, а амплітуда відображається яскравістю або кольором. На спектрограмі гітарної ноти на рис. 3.6 видно розвиток звучання: воно починається з різкої атаки і продовжується у вигляді гармонік, кратних за частотою основного тону 440 Гц.

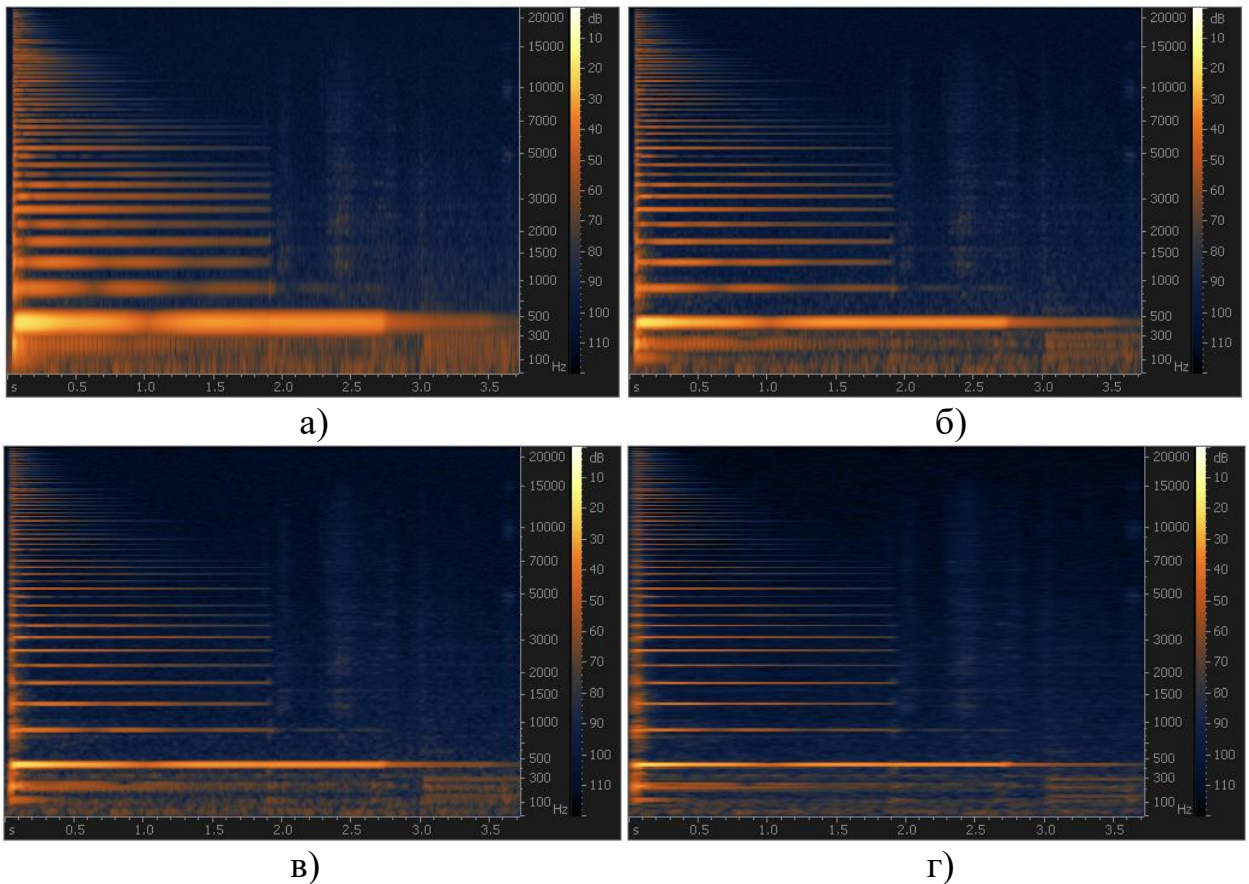


Рисунок 3.6 – Спектрограма гітарної ноти з різними розмірами вікна FFT (а – вікно 1024 відліків, б – 2048 відліків, в – 4096 відліків, г – 8192 відліків)

Видно, що верхні гармоніки мають меншу амплітуду і загасають швидше ніж нижні. Також на спектрограмі простежується шум запису – рівномірний фон темно-синього кольору. Справа показана шкала відповідності кольорів та рівнів сигналу (у дБ нижче нуля).

3.2 Дослідження спектрів мовних сигналів

Для дослідження спектрів мовних сигналів були завантажені звукозаписи дикторів в ефіру радіо, телебачення та Ютуб-каналів з мережі Інтернет. Задача – дослідити спектри мовних сигналів і встановити їх відповідність або відмінність від теоретичних оцінок, зроблених в розділі 2.

Слід розуміти, що аналізовані в даному розділі спектри вже пройшли еквалізацію, тобто звукорежисери довели їх до стану, придатного для мовлення. Наша задача – надати алгоритму автоматизованого керування логіки звукорежисера, що дозволить навіть непідготовленому користувачеві отримувати звукозаписи голосу, близькі за якістю до професійного рівня.

Зрозуміло, що автоматичні налаштування слід розглядати як рекомендацію для недосвідченого користувача і не виключає подальшого ручного тонкого налаштування АЧХ коригуючого еквалайзера.

Було завантажено ряд звукозаписів дикторів (рис.3.7) з наступних джерел:

- телеканал 1+1 (новини);
- Радіо Свобода (новини);
- Радіо НВ (ранкова програма);
- Ютуб канал «Давай запишемо».

Як видно з переліку, звукові відрізки взяті з різних за жанром та платформами програм. Задачею є виявити характерні особливості спектрів задля формування цільових функцій, до яких потрібно привести спектри будь-якого оброблюваного звукозапису.

Для порівняння були завантажені ряд необроблених звукозаписів невдалих дублів дикторів. Це зроблено для того, щоби дослідити відмінність спектрів «сирих» необроблених записів, і тих, які йдуть в ефір.

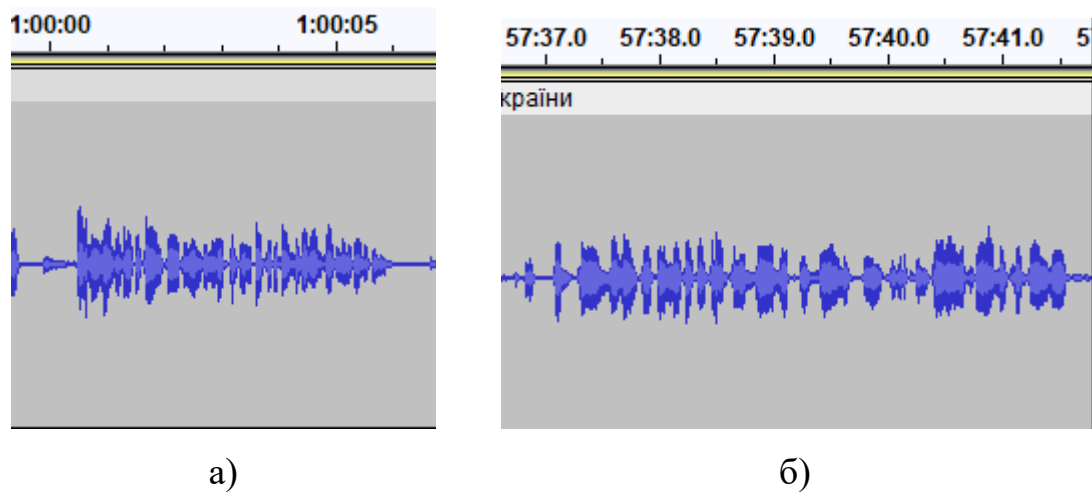


Рисунок 3.7 – Хвильоформи жіночого (а) і чоловічого (б) голосу дикторів з каналу 1+1 (передача новин ТСН)

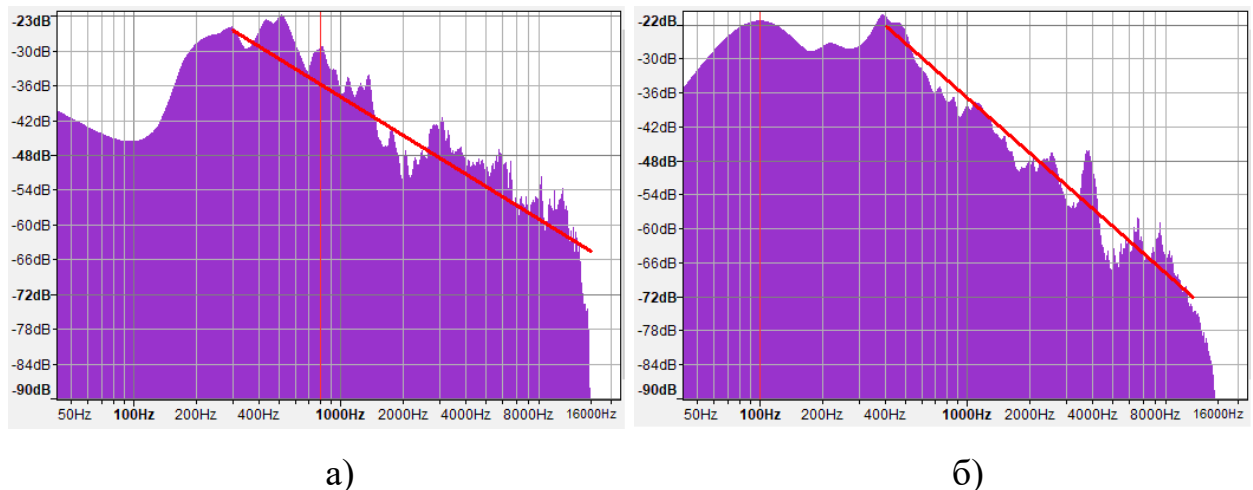


Рисунок 3.8 – Спектри жіночого (а) і чоловічого (б) голосу дикторів з каналу 1+1 (передача новин ТСН)

Для побудови спектрів (рис.3.8) використане швидке перетворення Фур'є з вікном Барлета тривалістю 1024 відліка при частоті дискретизації $F_s = 44,1$ кГц.

Цікаво відмітити, що нормалізація середнього рівня гучності сигналів у вітчизняному телевізійному мовленні відбувається по рівню -10 дБ. Очевидно, що звукорежисери залишають цей запас для трансляції реклами.

Спектр звукового супроводу телевізійної трансляції простирається майже до 15 кГц, що відповідає стандарту мовлення: 30 Гц ... 15 кГц.

Отримані спектри мають достатньо довгу тривалість реалізацій – більше 5 с. Тому вигляд спектру досить згладжений, без різких підйомів та спадів і деталізації. Розглянемо більш короточасні спектри тривалістю 1 с. На рис.3.9 показані реалізації жіночого голосу на Радіо Свобода, а на рис. 3.10 – відповідні їм спектри сигналів.

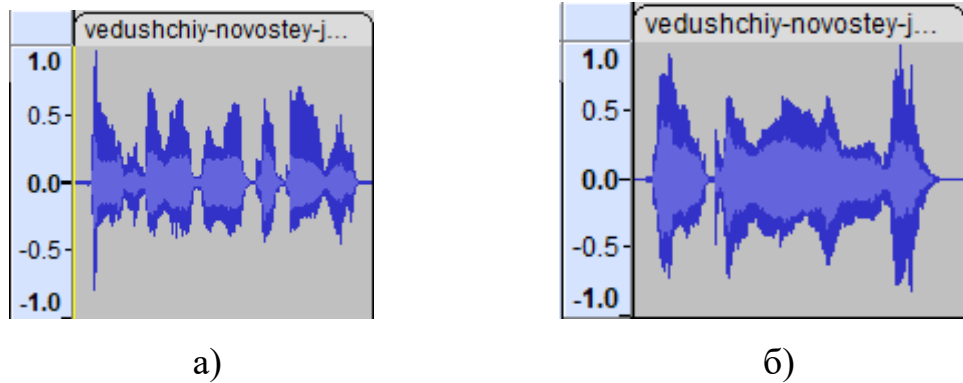


Рисунок 3.9 – Хвильоформи жіночого (а) і чоловічого (б) голосу дикторів з Радіо Свобода (передача новин)

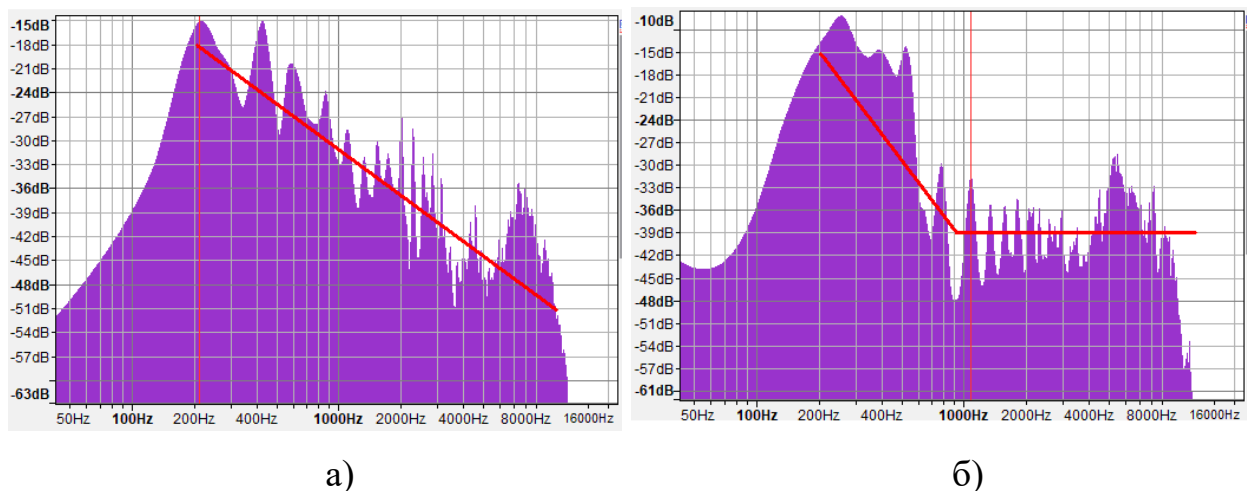


Рисунок 3.10 – Спектри жіночого (а) і чоловічого (б) голосу дикторів з Радіо Свобода (передача новин)

Аналізуючи графіки на рис.3.10 можна відмітити, що спадання середнього спектру жіночого голосу майже лінійне, а чоловічого – лінійне майже до частоти 1 кГц, а потім середній спектр можна вважати приблизно постійним. Тобто звукорежисери навмисно вирівнюють спектр СЧ та ВЧ компонент чоловічого голосу, щоби зробити його помітніше на фоні ВЧ

складових шуму. Шуми, найчастіше всього мають нормальний білий розподіл.

Для порівняння отримано «класичний» вигляд спектра сигналу звукозапису, зробленому на радянському радіо. Це розповідь фронтовички Другої Світової війни. На рис. 3.11, а показано хвилюформу сигналу даного звукозапису, а на рис. 3.11, б – середній спектр звукозапису, його тривалість 1 хв. 34 с.

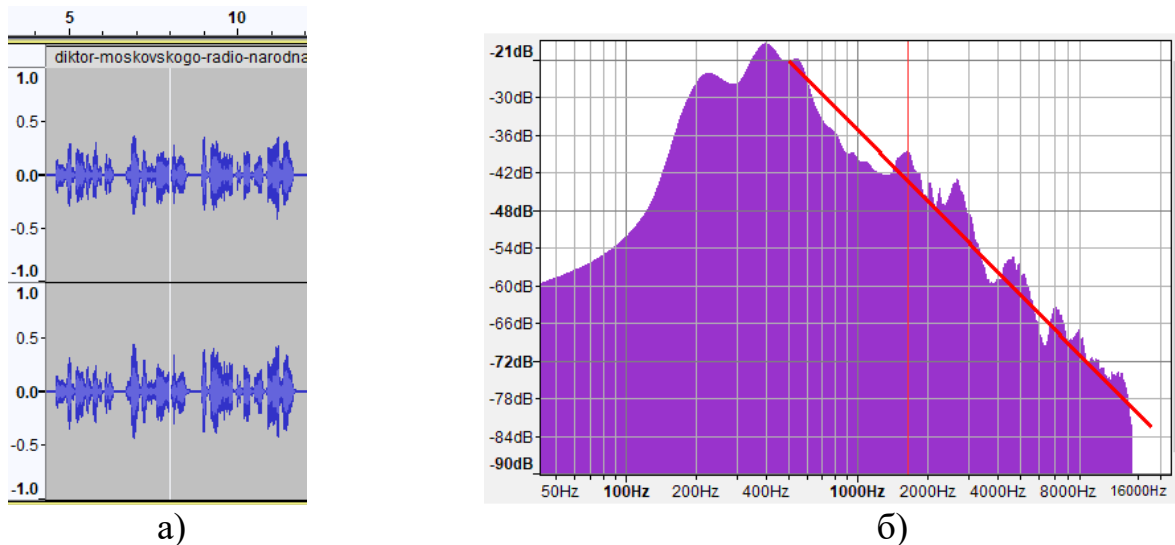


Рисунок 3.11 – Хвилюформа (а) і спектр (б) жіночого голосу з радянської радіостанції (розповідь фронтовички)

Спостерігаючи графіки на рис.3.11 можна відмітити класичний вигляд лінійно спадаючого спектру. Верхня частота обмежена стандартним значенням 15 кГц.

Для порівняння отримані ряд спектрів дикторів реклами. Даний жанр відрізняється від новостного тим, що потрібно мати поліпшену розбірливість мовної інформації в умовах маскуючої дії навколишніх акустичних шумів. Цього досягають одночасним застосуванням двох способів:

- стиснення динамічного діапазону звуку, зменшення його пік-фактору (за рахунок застосування компресорів та ліміттерів);
- підняття верхньої середини частотного спектру, починаючи з частоти 2 ... 3 кГц, що компенсує природний спад спектру мовного сигналу та робить помітніше дані частоти на фоні білого шуму з рівномірним спектром.

На рис.3.12, а показана типова хвилюформа дикторського мовного сигналу рекламного блоку, а на рис.3.12, б – його спектр.

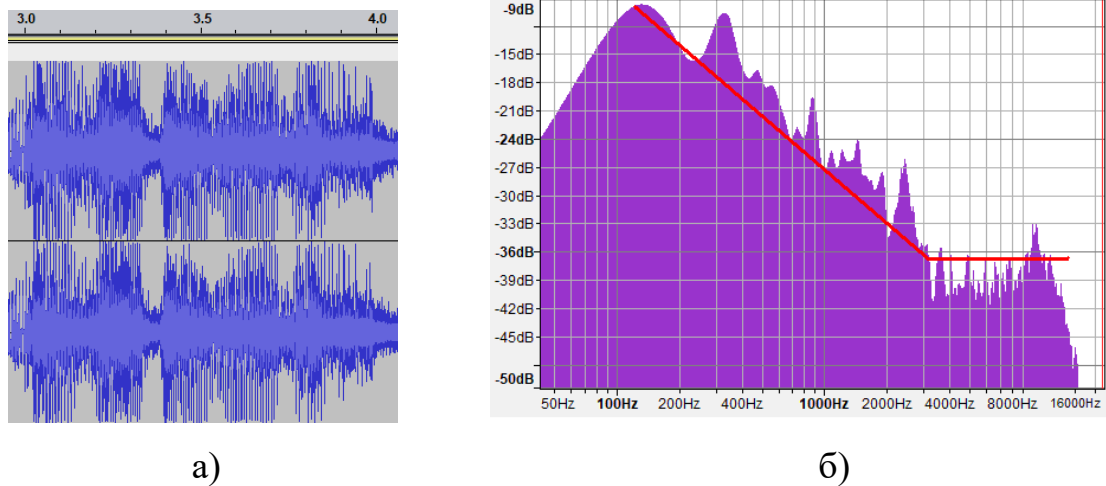


Рисунок 3.12 – Типова хвилюформа (а) і спектр (б) чоловічого голосу з рекламного блоку на радіо

Аналіз графіків на рис.3.12 а, б підтверджує зроблені вище висновки про стиснення динамічного діапазону та підняття верхньої середини спектру мови рекламної інформації.

Отримаємо оцінки швидкості спадання спектру мовних сигналів k_M для чоловічих і жіночих голосів. Результати внесемо в табл. 3.1. Розрахунок проведено згідно формули (3.3).

Таблиця 3.1 – Результати для розрахунку нахилів спектрів

Канал, програма	Голос, жін./чол.	Рисунок	k_M , дБ/окт.	$f_{нл}$, кГц
1+1, ТСН	Жіночий	Рис.3.8, а	7,5	–
1+1, ТСН	Чоловічий	Рис.3.8, б	9	<0,4
Радіо Свобода, новини	Жіночий	Рис.3.10, а	5,5	–
Радіо Свобода, новини	Чоловічий	Рис.3.10, б	10	>1,0
Рад. радіо	Жіночий	Рис.3.11, б	9,5	–
Радіо, реклама	Чоловічий	Рис.3.12, б	8	>3,0

$$k_M = \frac{S_M(f_2) - S_M(f_1)}{\log_2(f_2 / f_1)}, \text{ дБ/окт.} \quad (3.3)$$

Як впливає з аналізу оцінок, наведених в табл. 3.1, вони приблизно відповідають теоретичним даним, отриманим в розділі 2 ($k_M = 7,5$ дБ/окт.).

Швидкість спектрального спаду чоловічих голосів приблизно на 2...4 дБ більше, ніж у жіночих. Як середню оцінку можна вважати для жіночих голосів $k_M \approx 7$ дБ/окт., для чоловічих голосів – $k_M \approx 9$ дБ/окт.).

У разі необхідності збільшення розбірливості голосу доцільно робити підйом верхньої середини спектрів на частотах $f_{пл} > 1...3$ кГц таким чином, щоби обвідна спектру після корекції виявилась пласкою.

Підтвердимо зроблені висновки на практиці, оброблюючи записи голосів, зроблених в непередбачуваних умовах.

3.3 Експериментальне дослідження еквалізації мовних сигналів

Для того, щоби в'яснити можливості і обмеження запропонованого методу і його впливу на звучання, еквалізація виконувалась вручну з використанням аудіоредактора Audacity.

В якості вихідних звукозаписів були взяті записи голосів студентів групи ЕСТМ-22-1, які були зроблені ними в рамках курсу «Вступ у звукозапис». Такий підхід дозволив отримати максимальну різноманітність необробленого аудіо матеріалу, оскільки запис проводився в різних приміщеннях, на різну апаратуру, з індивідуальними тембральними особливостями, були присутні голоси як хлопців, так і дівчат. Всі аудіоматеріали мали формат WAV, 44100 Гц, 16 біт.

На першому етапі обчислювався спектр сигналу, параметри для даного обчислення показані на рис.3.13.



Рисунок 3.13 – Параметри для обчислення спектру

Тривалість вибірки 1024 і вікно Барлета вибрані з тих міркувань, щоби забезпечити достатню роздільну здатність і плавність побудови спектру без зайвої деталізації. На рис.3.14 показано один з прикладів спектру мовного сигналу.

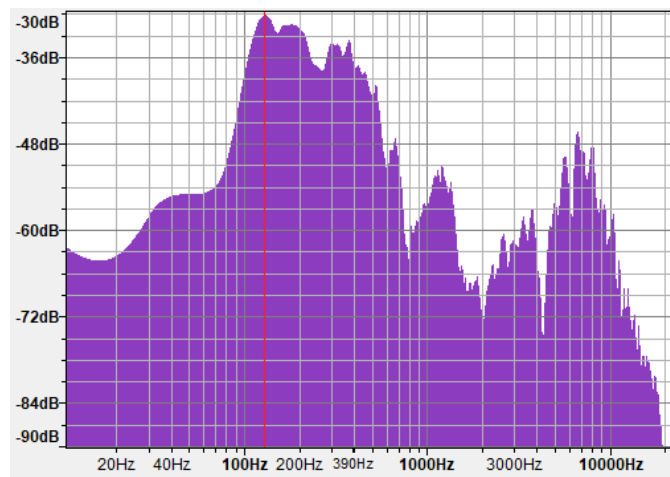


Рисунок 3.14 – Приклад спектру необробленого мовного сигналу

Візуально і на слух помітно провал в області середніх частот біля 2 кГц, а також зайвий підйом на частотах в околиці 7 кГц. Перший недолік проявляється як характерне звучання рокового вокалу, що не притаманно для дикторських голосів. Другий – як «шепелявлення», тобто виражені сибілянти – шиплячі та свистячі звуки в районі 6...8 кГц. Задачею еквалізації є усунення цих недоліків.

Для цього треба отримати оцінку обвідної спектру реального сигналу $S_p(f)$ і побудувати цільову обвідну спектру $S_u(f)$ після обробки (рис.3.15).

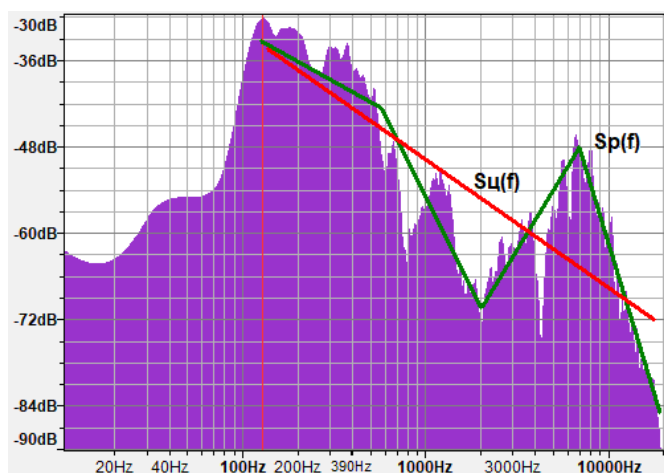


Рисунок 3.15 – Оцінка обвідної спектру реального сигналу $S_p(f)$ і цільова обвідна спектру $S_u(f)$ після обробки

Бажану криву АЧХ еквайзера корекції можна отримати як різницю між цільовою обвідною спектру $S_u(f)$ після обробки і оцінкою обвідної спектру реального сигналу $S_p(f)$

$$K_e(f) = S_u(f) - S_p(f), \text{ дБ.} \quad (3.4)$$

Реальна обробка не обов'язково повинна мати на виході точну цільову обвідну спектру $S_u(f)$, а лише давати певне наближення до неї. Тому реальну криву АЧХ еквайзера корекції слід отримати у вигляді

$$K_e(f) = k_e [S_u(f) - S_p(f)], \text{ дБ,} \quad (3.5)$$

де k_e – коефіцієнт, який обмежує ступінь вирівнювання АЧХ обробленого сигналу до цільової обвідної спектру, $k_e=0\dots 1$, значення $k_e=0$ відповідає відсутності еквалізації, $k_e=1$ означає повну корекцію АЧХ.

На рис.3.16 показано вид АЧХ еквайзера корекції $K_e(f)$, отриманої для спектру необробленого сигналу на рис.3.15 згідно виразу (3.5) при $k_e=1$.

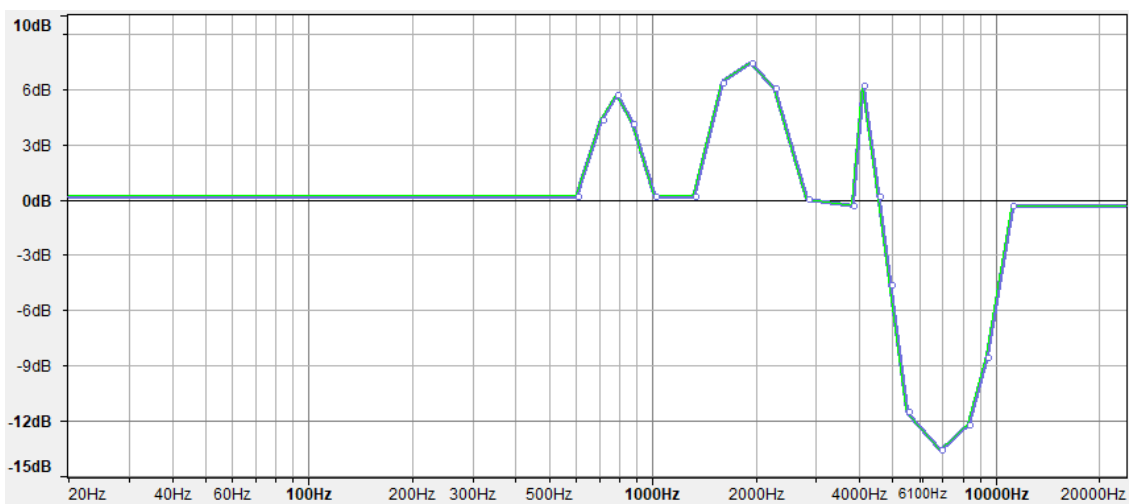


Рисунок 3.16 – АЧХ еквайзера корекції $K_e(f)$, отриманої для спектру необробленого сигналу на рис.3.15 згідно (3.5) при $k_e=1$

На рис.3.17 наведено результат корекції АЧХ сигналу на рис.3.15 кривою еквайзера на рис.3.16 при $k_e=1$.

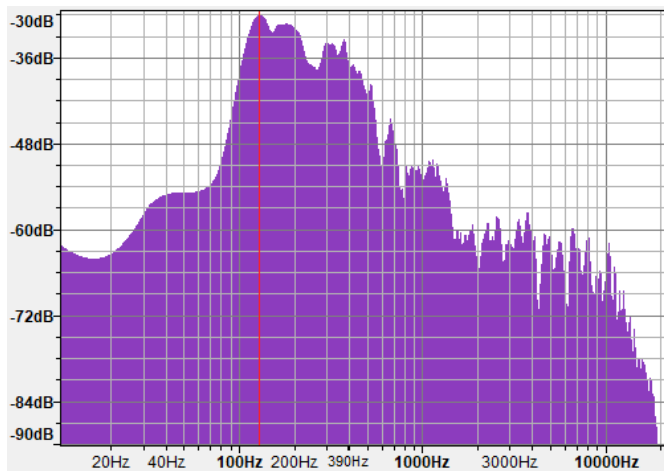


Рисунок 3.17 – Результат корекції АЧХ мовного сигналу

Після корекції візуально і на слух помітно ліквідацію провалу в області середніх частот біля 2 кГц, а також зайвого підйому в районі сибілянтів мови на частотах 6...8 кГц. Звучання стало набагато комфортнішим для прослуховування, що проявляється як відсутність стомлюваності слухового апарату при тривалому прослуховуванні. Отриманий запис за звучанням наближається до якості телевізійного і радіомовлення, незважаючи на те, що оригінальна звукова доріжка записана на диктофон смартфона.

Оброблені ще ряд записів, результати експериментів показані на рис.3.18 – рис.3.23.

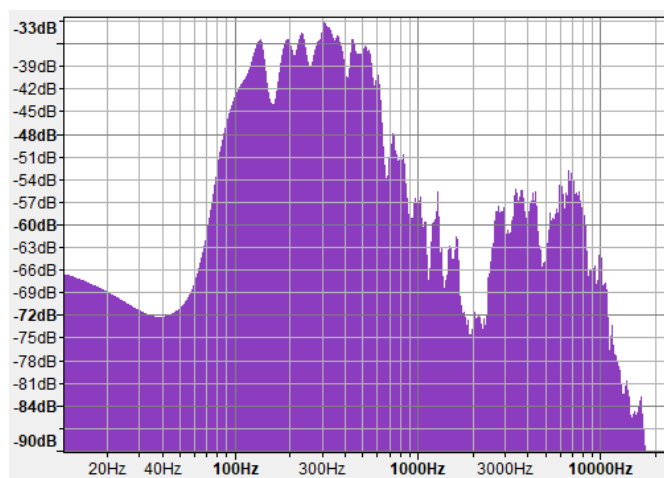


Рисунок 3.18 – Спектр необробленого мовного сигналу (приклад 2)

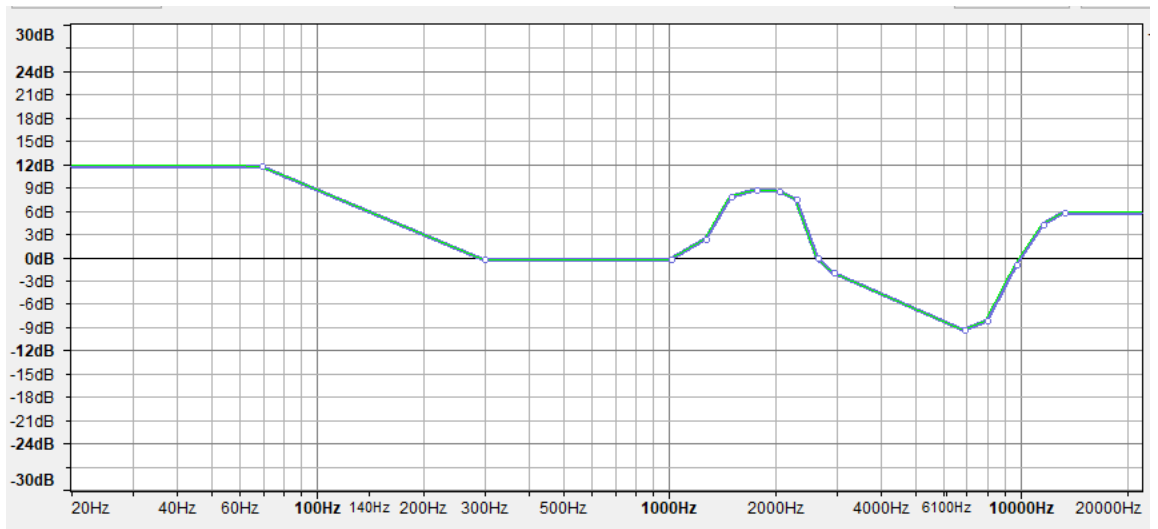


Рисунок 3.19 – АЧХ еквайзера корекції для необробленого сигналу
(приклад 2)

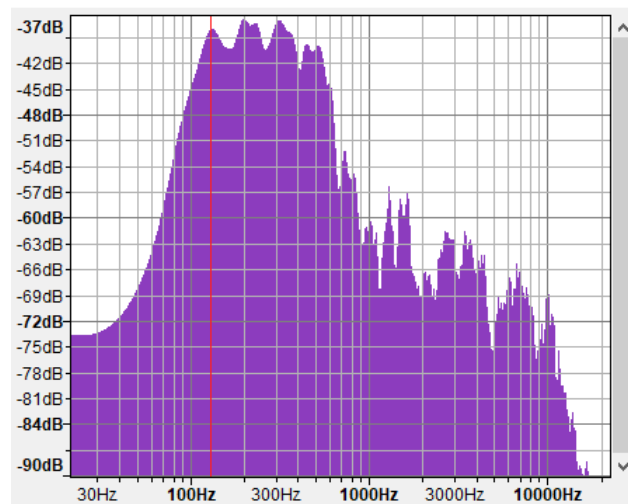


Рисунок 3.20 – Результат корекції АЧХ мовного сигналу (приклад 2)

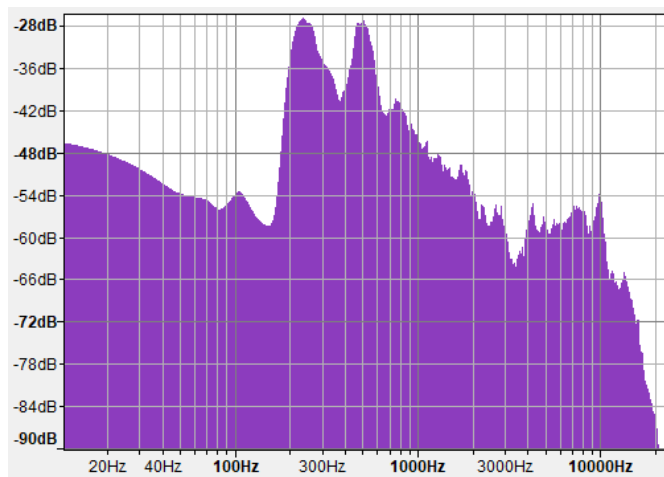


Рисунок 3.21 – Спектр необробленого мовного сигналу (приклад 3)

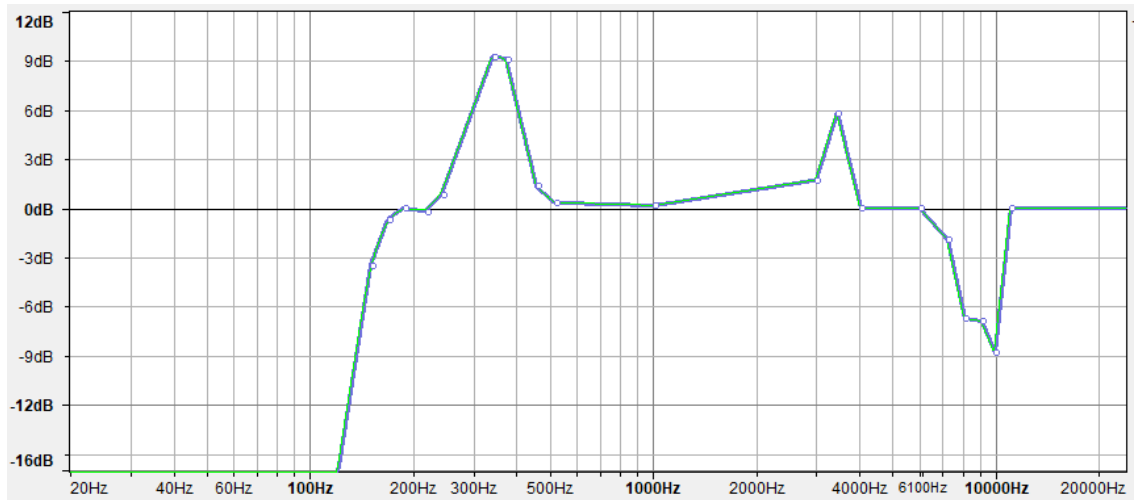


Рисунок 3.22 – АЧХ еквайзера корекції для необробленого сигналу (приклад 3)

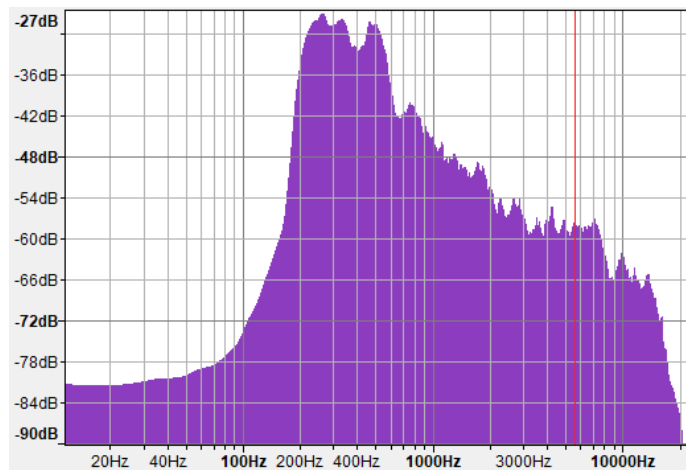


Рисунок 3.23 – Результат корекції АЧХ мовного сигналу (приклад 3)

У прикладі 2 використано запис голосу хлопця, у прикладі 3 – дівчини. Це видно з частоти основного тону 150 Гц у хлопця (приклад 2) і 250 Гц у дівчини (приклад 3). Відповідно відрізняється обробка в області НЧ: у прикладі 2 використано підйом НЧ до 12 дБ для придання голосу додаткової мужності, у прикладі 3 – обрізання НЧ зі швидкістю 20 дБ/окт. для зменшення непотрібних призвуків у вигляді поп-ефекту.

Еквалізація проводилася за виразом (3.5). Експериментально встановлено, що найкращий результат виходить при коефіцієнті корекції $k_e=0,7\dots0,8$. Це дозволяє вирівняти загальний вигляд спектру до лінійно

спадаючого зі швидкістю 7...8 дБ/окт. при збереженні індивідуальних особливостей голосу. В цілому завдяки еквалізації вдається отримати запис наближений за звучанням до якості телевізійного і радіомовлення.

3.4 Висновки по розділу 3

Експериментально виявлено характерні особливості спектрів мовних сигналів з метою формування цільових функцій, до яких потрібно привести спектри будь-якого оброблюваного звукозапису. Для дослідження були завантажені звукозаписи дикторів в ефіру радіо, телебачення та Ютуб-каналів з мережі Інтернет. Задачею було надати алгоритму автоматизованого керування логіки звукорежисера, що дозволить навіть непідготовленому користувачеві отримувати звукозаписи голосу, близькі за якістю до професійного рівня.

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

Швидкість спектрального спаду чоловічих голосів приблизно на 2...4 дБ більше, ніж у жіночих. Як середню оцінку можна вважати для жіночих голосів 7 дБ/окт., для чоловічих – 9 дБ/окт.

У разі необхідності збільшення розбірливості голосу доцільно робити підйом верхньої середини спектрів на частотах більше 1...3 кГц таким чином, щоби обвідна спектру після корекції виявилась пласкою.

Дані висновки підтверджено на практиці, шляхом обробки записів голосів, зроблених в непідготованих умовах. В якості вихідних звукозаписів були взяті записи голосів студентів групи ЕСТМ-22-1, які були зроблені ними в рамках курсу «Вступ у звукозапис». Такий підхід дозволив отримати максимальну різноманітність необробленого аудіо матеріалу, оскільки запис проводився в різних приміщеннях, на різну апаратуру, з індивідуальними тембральними особливостями, були присутні голоси як хлопців, так і дівчат. Всі аудіоматеріали мали формат WAV, 44100 Гц, 16 біт.

На першому етапі обчислювався спектр сигналу. Тривалість вибірки 1024 і вікно Барлета вибрані з тих міркувань, щоби забезпечити достатню роздільну здатність і плавність побудови спектру без зайвої деталізації. У необроблених сигналів і візуально і на слух були спектральні вади, які полягають у підкресленні або провалі деяких смуг частот.

Бажана крива АЧХ еквалайзера корекції отримувалася як різниця між цільовою обвідною спектру після обробки і оцінкою обвідної спектру необробленого сигналу.

Реальна обробка не обов'язково повинна мати на виході точну цільову обвідну спектру, а лише давати певне наближення до неї. Тому введений коефіцієнт, який обмежує ступінь наближення АЧХ обробленого сигналу до цільової обвідної спектру: 0 відповідає відсутності еквалізації, 1 означає максимальну корекцію АЧХ.

Експериментально встановлено, що найкращий результат виходить при коефіцієнті корекції 0,7...0,8. Це дозволяє вирівняти загальний вигляд спектру до лінійно спадаючого зі швидкістю 7...8 дБ/окт. при збереженні індивідуальних особливостей голосу. Звучання стає набагато комфортнішим для прослуховування, що проявляється як відсутність стомлюваності слухового апарату при тривалому прослуховуванні. В цілому завдяки еквалізації вдається отримати запис наближений за звучанням до якості телевізійного і радіомовлення, незважаючи на те, що оригінальні звукові доріжки записані на диктофон смартфона.

ВИСНОВКИ

Запис голосу диктора – дуже важливий етап аудіовиробництва. Після того, як запис диктора виконано, настає етап редагування.

Частотна корекція або як кажуть звукорежисери – еквалізація – один з найпопулярніших ефектів для обробки голосу. Також, на жаль, це один із ефектів, яким часто зловживають. При неправильній еквалізації, кінцевий результат може звучати навіть гірше, ніж до використання.

Мета роботи – розробити методику автоматизованої еквалізації, яка дасть можливість навіть початківцю привести аудіозапис до поширених нині у мовленні і звичних для слухачів стандартів АЧХ.

В результаті теоретичного обґрунтування принципів еквалізації мовних сигналів показано, що основним показником якості передачі у вирішуваний задачі є натуральність звучання, що забезпечується ідентичністю тембру джерела і відтворюваного звуку.

Розглянуто структурну схему наскрізного тракту запису і відтворення звуку. Отримано вираз для АЧХ еквалайзера корекції. Не зважаючи на уявну простоту виразу, проблема полягає в тому, що заздалегідь невідомий вигляд АЧХ приміщення і мікрофону. Навіть якщо застосовується мікрофон з відомою АЧХ, реальний вигляд АЧХ відрізнятиметься від паспортних даних за рахунок невідомого розміщення диктора відносно мікрофона. АЧХ звучання приміщення теж залежить від місць розташування мікрофону і джерела звуку (диктора). Тому треба вирішити некоректну в математичному сенсі задачу визначення АЧХ еквалайзера в умовах апріорної невизначеності функцій АЧХ приміщення і мікрофону.

Єдиним «випробувальний» сигналом, що є в системі, це мова диктора. Ми можемо користуватися лише оціночним значенням спектра голосу диктора, яку і будемо використовувати як основу для налаштування еквалайзера.

Проведено теоретичні оцінки середньої швидкості спадання спектру мовних сигналів. Для цього було апроксимовано пласку частину обвідної спектру та обвідної формант лінійною функцією, враховуючи, що графіки побудовані в логарифмічному масштабі. Виводячи середнє значення між швидкостями спадання спектру мови та формант мови отримано значення 7,5 дБ/окт.

Слід очікувати, що при такій швидкості спаду мова буде мати натуральне звучання. Дана швидкість спаду має бути для частот від 200...300 Гц і до верхніх частот звукового діапазону. Область нижніх частот, нижче 200...300 Гц доцільно не чіпати і не оброблювати еквайзером.

Для уточнення отриманих результатів проведено експериментальне дослідження аудіо записів реальних ефірів на радіо, телебаченні і в Ютубі. Аналізовані спектри вже пройшли еквалізацію звукорежисерами. Наша задача – надати алгоритму автоматизованої еквалізації логіки звукорежисера.

Звукові відрізки взяті з різних за жанром та платформами програм. звукові відрізки взяті з різних за жанром та платформами програм. Для побудови спектрів використане швидке перетворення Фур'є з вікном Барлета тривалістю 1024 відліка при частоті дискретизації 44,1 кГц. Тривалість вибірки 1024 і вікно Барлета вибрані з тих міркувань, щоби забезпечити достатню роздільну здатність і плавність побудови спектру без зайвої деталізації.

Всі досліджені спектри за своїм виглядом можна умовно поділити на два типи: класичний лінійно спадаючий (або новостний) та рекламний. Рекламний жанр відрізняється від новостного тим, що потрібно мати поліпшену розбірливість мовної інформації в умовах маскуючої дії навколишніх акустичних шумів. Для збільшення розбірливості звукорежисери роблять підйом верхньої середини спектру на частотах більше 1...3 кГц таким чином, щоби обвідна спектру після корекції виявилась пласкою.

Експериментально встановлено, що швидкість спектрального спаду чоловічих голосів приблизно на 2...4 дБ більше, ніж у жіночих. Як середню

оцінку можна вважати для жіночих голосів 7 дБ/окт., для чоловічих – 9 дБ/окт.

Дані висновки підтверджено на практиці, шляхом обробки записів голосів, зроблених в непередбачених умовах. В якості вихідних звукозаписів були взяті записи голосів студентів групи ЕСТМ-22-1, які були зроблені ними в рамках курсу «Вступ у звукозапис». Такий підхід дозволив отримати максимальну різноманітність необробленого аудіо матеріалу, оскільки запис проводився в різних приміщеннях, на різну апаратуру, з індивідуальними тембральними особливостями, були присутні голоси як хлопців, так і дівчат. Всі аудіоматеріали мали формат WAV, 44100 Гц, 16 біт.

Бажана крива АЧХ еквалайзера корекції отримувалася як різниця між цільовою обвідною спектру після обробки і оцінкою обвідної спектру необробленого сигналу.

Реальна обробка не обов'язково повинна мати на виході точну цільову обвідну спектру, а лише давати певне наближення до неї. Тому введений коефіцієнт, який обмежує ступінь наближення АЧХ обробленого сигналу до цільової обвідної спектру: 0 відповідає відсутності еквалізації, 1 означає максимальну корекцію АЧХ.

Експериментально встановлено, що найкращий результат виходить при коефіцієнті корекції 0,7...0,8. Це дозволяє вирівняти загальний вигляд спектру до лінійно спадаючого зі швидкістю 7...8 дБ/окт. при збереженні індивідуальних особливостей голосу. Звучання стає набагато комфортнішим для прослуховування, що проявляється як відсутність стомлюваності слухового апарату при тривалому прослуховуванні. В цілому завдяки еквалізації вдається отримати запис наближений за звучанням до якості телевізійного і радіомовлення, незважаючи на те, що оригінальні звукові доріжки записані на диктофон смартфона.

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

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Психоакустические аспекты восприятия речи. Механизмы деятельности мозга / Под. ред. Н. П. Бехтеревой. — М. Наука. 1988. — 504. с.
2. Фланаган Дж. Анализ, синтез и восприятие речи: Пер. с англ./ Под ред. А. А. Пирогова. — М. Связь. 1968. —396 с.
3. Вокодерная телефония. Методы и проблемы. Под ред. А. А. Пирогова — М. Связь. 1974. — 536 с.
4. Назаров М. В., Прохоров Ю. Н. Методы цифровой обработки и передачи речевых сигналов. — М. Радио и связь. 1985. — 176 с.
5. Михайлов В. Г., Златоустова Л. В. Измерение параметров речи /Под ред. М. А. Сапожкова. — М. Радио и связь. 1987. — 168 с.
6. Вемян Г. В. Передача речи по цепям электросвязи. — М. Радио и связь. — 1985. — 272 с.
7. Зубов А.Ю. ОБЕРТОН. М.: Большая Российская энциклопедия. — Т. 23. — 2013.
8. ANSI S1.1-1994 (ASA 111-1994) Revision of ANSI S1.1-1960 (R1976). Reaffirmed by ANSI March 25, 2004. Includes 2005 Erratum.
9. Jones M.R., Fay R.R., Popper A.N. Music Perception. New York: Springer, 2011. — 364 p.
10. Klingbeil M.K. Spectral Analysis, Editing, and Resynthesis: Methods and Applications. New York: Columbia University, 2009. — 468 p.
11. ТЕМБР. Большая Российская энциклопедия. Том 32. Москва. 2016.
12. Clarkson MG, Clifton RK, Perris EE. Infant timbre perception: Discrimination of spectral envelopes. Perception & Psychophysics. 1988; 43(1). — p. 15-20.
13. Алдошина И.А., Приттс Р. Музыкальная акустика. Учебник. СПб.: Композитор, 2006. — 720 с.
14. Радиовещание и электроакустика / С.И. Алябьев, А.В. Выходец, Р. Гермер и др.; под ред. Ю.А. Ковалгина. — М.: Радио и связь, 2000. — 792 с.

15. В.Н. Олейников, О.В. Зубков, В.М. Карташов, И.В. Коротцев, 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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 ХХІ столітті». Зб. матеріалів форуму. Т. 3. – Харків: ХНУРЕ. 2021. – с. 119 – 120.

32. Покровский Н.Б. Расчет и измерение разборчивости речи. – М., Связьиздат, 1962. – 390с.

33. 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.