МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ВАСИЛЯ СТУСА

БОДОВСЬКИЙ ВЛАДИСЛАВ ДМИТРОВИЧ

Допускається до захисту
Завідувач кафедри
Інформаційних технологій,
к. т. н., доцент
_____О. В. Зелінська
«_____»_____2024

ДОСЛІДЖЕННЯ МАТЕМАТИЧНИХ МОДЕЛЕЙ ТА МЕТОДІВ АНАЛІЗУ АУДІОСИГНАЛУ

Спеціальність 122 «Комп`ютерні науки»

Кваліфікаційни (магістерська) робота

Науковий керівник:
Є.Є. Федоров, д.т.н., професор,
професор кафедри
інформаційних технологій

_____
(підпис)

Оцінка: _____/_____/_____
(балів за шкалою ЄКТС/за національною шкалою)
Голова ЕК:_____
(підпис)

Вінниця — 2024

Анотація

Кваліфікаційна (магістерська) дипломна робота складається з 70 сторінок формату А4, на яких є 38 рисунків, список використаних джерел містить 30 найменувань.

У кваліфікаційній дипломній роботі проаналізовано принцип роботи математичних моделей та методів аналізу аудіосигналу.

Сформульовано мету досліджень – удосконалення та спрощення процесу аналізу аудіосигналу шляхом розробки власних модулей поверхневого та глибокого звукового аналізу.

Запропоновано метод вирішення питання про аналіз аудіосигналу на предмет присутності на ньому штучно згенерованого людського голосу. Проведено поділ методів генерації, що визначає актуальну глибину аналізу. Висвітлено принцип математичних моделей та методів обчислення ключових характеристик аудіосигналу. Проведено аналіз змішаного методу генерації штучного голосу на основі природного. На основі отриманих даних про алгоритми конструювання синтетичного голосу таким методом було визначено його можливі слабкості та визначено ключові характеристики, потрібні для аналізу запису, що потенційно може містити в собі такий штучний голос. Розроблено модулі обчислення ключових характеристик, таких як: Мел-кепстральний коефіцієнт, спектрограма, діаграма спектрального центроїду аудіосигналу, діаграма спектрального спаду, діаграма спектрального контрасту, хромограма.

Аргументовано актуальність та доцільність обраних програмних записів для написання цього програмного додатку.

Розроблено програмний додаток для аналізу аудіосигналу на предмет присутності у ньому синтетичного людського голосу.

Отримані в магістерській дипломній роботі результати можна використати для побудови високопродуктивних систем аналізу аудіосигналу.

Ключові слова: звук, аудіосигнал, аналіз, методи аналізу аудіосигналу, аудіо дані, штучний голос.

Abstract

The qualifying (master's) diploma thesis consists of 70 pages of A4 format, on which there are 38 figures, the list of used sources contains 30 titles.

In the qualification thesis the working principle of mathematical models and methods of audio signal analysis been analyzed.

The purpose of the research is formulated - to improve and simplify the process of audio signal analysis by developing own modules of surface and deep sound analysis.

A method of solving the problem of analyzing an audio signal for the presence of an artificially generated human voice is proposed. The generation methods are divided, which determines the actual depth of the analysis. The principle of mathematical models and methods of calculating the key characteristics of an audio signal is highlighted. An analysis of the mixed method of generating an artificial voice based on a natural one was carried out. On the basis of the obtained data on the algorithms for the construction of a synthetic voice by this method, its possible weaknesses were determined and the key characteristics necessary for the analysis of a recording that could potentially contain such an artificial voice were determined. Modules have been developed for calculating key characteristics, such as: Mel-cepstral coefficient, spectrogram, diagram of the spectral centroid of the audio signal, diagram of spectral decay, diagram of spectral contrast, chromogram.

The relevance and expediency of the selected software entries for writing this software application are argued.

A software application has been developed for analyzing an audio signal for the presence of a synthetic human voice in it.

The results obtained in the master's thesis can be used to build high-performance audio signal analysis systems.

Keywords: sound, audio signal, analysis, audio signal analysis methods, audio data, artificial voice.

Зміст

ВСТУП

Сучасний розвиток інформаційних технологій відкриває безмежні можливості для застосування штучного інтелекту в різних сферах життя. Однією з найбільш перспективних галузей є програмна генерація та програмний аналіз звуку, який стає надзвичайно важливим у сучасному світі, де розпізнавання та ідентифікація аудіозаписів стають важливими безпековими завданнями.

Штучний голос людини, створений з використанням синтезу звуку, відкриває нові перспективи в комунікаціях та взаємодії з комп'ютерними системами, а також дозволяє виконувати безліч різноманітних робіт, прикладами яких можуть бути:

- Генерація штучного голосу людини для озвучування та дубляжу персонажів фільмів, мультфільмів, та інших видів аудіовізуального мистецтва;
- Інтеграція технології голосового помічника для зручності користування іншими технологіями;
- Оптимізація та покращення якості навчального процесу з вивчення іноземних мов;
- Генерація штучного голосу для анонімного ведення розмови у мережі, що може стати дуже корисною технологією для відомих людей, а також стрімерів;
- За допомогою технології генерації голосу є можливим створення вокалу для пісні за текстом, який не встиг реалізувати за різних причин вокаліст того чи іншого гурту;
- Тощо.

Проте, разом із зростанням зацікавленості в цій області, з'являється необхідність в розробці та вдосконаленні методів програмного аналізу звуку для ефективної ідентифікації та аналізу аудіосигналу на предмет присутності на ньому штучного голосу. Адже з розвитком технології генерації голосу стають доступними нові схеми шахрайства, що опираються на копіювання голосу певної людини, що робить можливим наступні речі:

– Обхід голосової ідентифікації в різних банківських системах;
– Спрощення схем телефонного шахрайства;
– Зменшення надійності голосових паролів;
– Тощо.

Ця робота присвячена дослідженню математичних моделей та методів програмного аналізу аудіосигналу з метою ідентифікації штучного голосу людини на аудіозаписах, а також покращенню цих методів. Дослідження у цій області є актуальним як для розвитку систем нейромереж та штучного інтелекту, так і для забезпечення безпеки перебування в онлайн-середовищі та реальному світі. Результати досліджень можуть знайти практичне застосування як у сферах біометричної ідентифікації, аутентифікації користувачів так і у сферах управління доступом до інформаційних ресурсів. Подальшого розвитку отримав блок ідентифікації ключових параметрів аудіосигналу, що може бути використаний іншими системами аналізу аудіосигналу для ідентифікації синтетичного голосу, що був створений стандартним або змішаним методом.

У пояснювальній записці до кваліфікаційної (магістерської) дипломної роботи було розглянуто 4 розділи та було використане ? літературне джерело.

РОЗДІЛ 1
ПОСТАНОВКА ЗАДАЧІ І ОГЛЯД АНАЛОГІВ ІСНУЮЧИХ ДОДАТКІВ ЗА ДАНОЮ ТЕМАТИКОЮ

1.1 Опис актуальності дослідження математичних моделей та методів аналізу аудіосигналу

Аналіз аудіосигналів та розробка математичних моделей та методів для його обробки є актуальною галуззю досліджень у сучасному світі. Існує безліч ситуацій, коли такий аналіз є необхідним і корисним, і його результати можуть бути застосовані у різних сферах.

Однією з основних причин актуальності досліджень з аналізу аудіосигналу є зростаючий обсяг аудіоінформації, що є доступною для обробки. Завдяки швидкому розвитку цифрових медіа технологій, люди стають виробниками і споживачами великих обсягів аудіоданих. Застосуванням математичних моделей та методів аналізу аудіосигналу можна ефективно обробляти цю інформацію, витягати з неї корисні дані і отримувати нові знання [1].

Аналіз аудіосигналу має широкий спектр застосувань. Наприклад, у сфері музики та звукозапису математичні моделі та методи аналізу аудіосигналу допомагають виявляти музичні структури, використовувати автоматичне розпізнавання музичних жанрів та інструментів, аналізувати та виправляти звукові дефекти. У сфері мовлення такий аналіз використовується для розпізнавання та синтезу мови, виявлення емоцій у голосі, аналізу мовного поведінки [2].

Дослідження математичних моделей та методів аналізу аудіосигналу також мають важливе значення у сферах безпеки і захисту. Вони можуть використовуватись для аналізу звукових сигналів з метою виявлення небезпечних звуків, таких як вибухи, вистріли зі зброї або звуків тривоги. Такий аналіз може бути корисним у системах відеоспостереження, системах контролю звуків у громадських приміщеннях або системах безпеки на транспорті [3].

Також з розвитком технологій штучного інтелекту та нейромереж актуальною стає проблема у відрізненні голосу живої людини від синтетичного змодульованого голосу [4].

1.2 Постановка задачі

Об’єктом дослідження в рамках магістерської атестаційної роботи є процес аналізу аудіосигналу. Предметом дослідження є методи аналізу звукової доріжки для визначення її ключових параметрів, що дозволять відрізнити натуральне мовлення від синтетичного. Метою даної роботи є дослідження методів визначення Мел-кепстрального коефіцієнту, спектрального центроїду, контрасту, спаду та інших ключових параметрів. Також метою є розробка ефективних алгоритмів, які дозволять визначати особливості та характеристики штучного голосу з точністю, необхідною для подальшої його ідентифікації, та розробка програмного забезпечення для демонстрації роботи алгоритмів. Для досягнення мети, необхідно досліджувати наступні питання:

- Проаналізувати існуючі методи аналізу аудіоданих;
- Обґрунтувати мету дослідження та покращення алгоритмів аналізу аудіоданих;
- Розробити вдосконалену модель аналізу аудіоданих;
- Застосувати вдосконалену модель аналізу аудіоданих разом з існуючими методами задля відображення його переваг над класичної моделлю.

1.2.1 Обґрунтування мети та формування вимог до удосконалення методу аналізу аудіоданих

Якісно та лаконічно організований блок аналізу даних дозволяє не тільки знизити показник необхідного і достатнього часу для обробки аудіоданих, але й підвищити точність визначення потрібних коефіцієнтів, що знизить шанс помилки, та зменшить кількість невизначених ситуацій, що однозначно покращить роботу системи. Саме для цього необхідно постійно покращувати математичні моделі та методи аналізу аудіоданих.

Покращення алгоритмів аналізу аудіоданих потрібна для того, щоб:

– Зменшити витрати часу для обчислення одного випадку;
– Зменшити витрати обчислювальних ресурсів для зниження навантаження на систему;
– Збільшення точності визначення наявності синтетичного голосу на записі.

1.2.2 Вимоги до методу аналізу аудіоданих

Спираючись на результати аналізу існуючих методів розв`язання подібних задач та проблем, вимоги до удосконалення наступні:

– вдосконалення методу аналізу за допомогою удосконалення обрахунку ключових параметрів, що дозволяють відрізнити природне мовлення від синтетичного;

– застосування вдосконаленої моделі та обчислення Мел-кепстрального коефіцієнту, спектрограми, спектрального центроїду, спектрального контрасту, спектрального спаду та інших ключових характеристик;

– використання методу не повинно займати багато часу.

1.3 Огляд аналогів

Наразі існує кілька програмних рішень, які можуть допомогти у розрізненні природного та синтетичного голосу. Одним з найпопулярніших пакетів програмного забезпечення для цієї мети є Praat.

Praat — це безкоштовний комп’ютерний пакет програм для аналізу мовлення у фонетиці. Його розробили та продовжують розвивати Пол Боерсма та Девід Венінк з Амстердамського університету. Він може працювати на широкому спектрі операційних систем, включаючи різні версії Unix, Linux, Mac і Microsoft Windows (2000, XP, Vista, 7, 8, 10). Програма підтримує синтез мовлення, в тому числі артикуляційний [5]. Це програмне забезпечення має потужний набір інструментів в вигляді багатьох функцій, використовувати які можна написавши відповідний

скрипт. Але такий підхід має свої недоліки у вигляді високого порогу входження для звичайного користувача [6].

Praat надає користувачам можливість виконувати різноманітні завдання, пов'язані з аналізом звуку. Основні функції Praat включають:

Аналіз формантів: Praat дозволяє вимірювати форманти — резонансні піки, які спостерігаються в спектрі голосу та допомагають ідентифікувати голосові звуки.

Спектральний аналіз: Програма дає змогу вивчати спектральні характеристики звуків, зокрема ширину смуги, інтенсивність та форму спектра.

Аналіз формантних траекторій: Praat дозволяє вивчати зміну формантів протягом часу, що є важливим для аналізу мовлення та мовних звуків.

Синтез мови: За допомогою Praat можна створювати штучний голос та синтезувати мову на основі різних параметрів.

Транскрипція: Praat надає можливість транскрибувати аудіозаписи, перетворюючи мовлення на письмову форму.

Praat є популярним інструментом у наукових дослідженнях в області фонетики, лінгвістики та спілкування загалом. Він надає багато можливостей для аналізу та обробки голосу та звуку, дозволяючи вченим, лінгвістам та іншим спеціалістам вивчати різні аспекти мовлення.

Підбиваючи підсумки, можна сказати, що перевагами Praat є:

1. Розширений аналіз: Praat надає широкий спектр можливостей для аналізу голосу і звуку. Він дозволяє вимірювати форманти, вивчати спектральні характеристики, аналізувати формантні траекторії та багато іншого. Це робить його важливим інструментом для дослідження фонетики та мови.
2. Гнучкість: Praat дає користувачам велику гнучкість в роботі. Ви можете виконувати різні завдання з аналізу звуку, використовуючи широкий набір параметрів та налаштувань. Це дозволяє вам адаптувати інструмент до своїх потреб і вимог дослідження.

3. Безкоштовний доступ: Praat є вільно доступним програмним забезпеченням, що означає, що ви можете завантажити його безкоштовно та використовувати для своїх досліджень або проєктів.

Недоліками ж в свою чергу є:

1. Складність використання: Для новачків може знадобитися час і стеження посібника або навчального курсу, щоб зрозуміти, як користуватися Praat. Інтерфейс може виглядати складним для тих, хто не має попереднього досвіду в аналізі голосу.
2. Обмежена візуалізація: Інтерфейс Praat має деякі обмеження візуалізації даних. Іноді користувачам може бути складно чітко представити та відображати результати аналізу.
3. Обмежена підтримка форматів файлів: Praat підтримує обмежену кількість форматів аудіофайлів, що може обмежити користувачів у роботі зі своїми власними даними. [7].

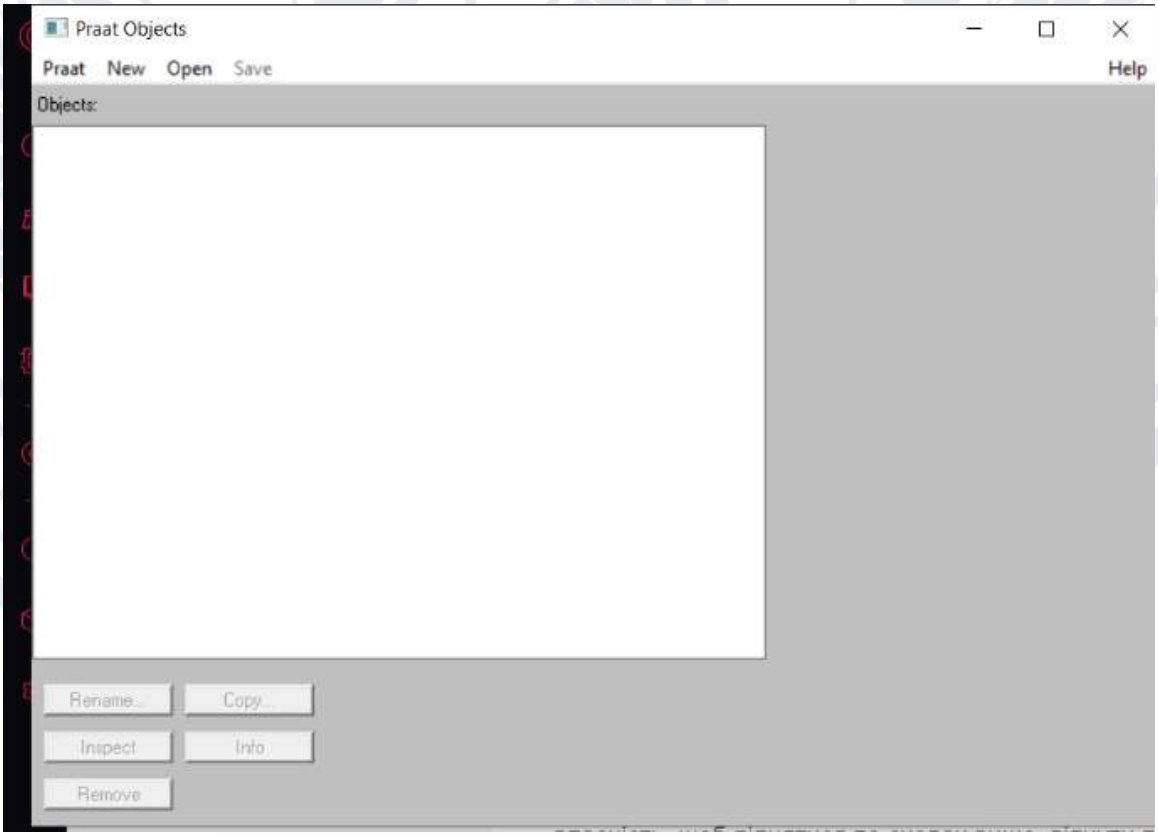


Рисунок 1.1 - Сторінка вводу даних

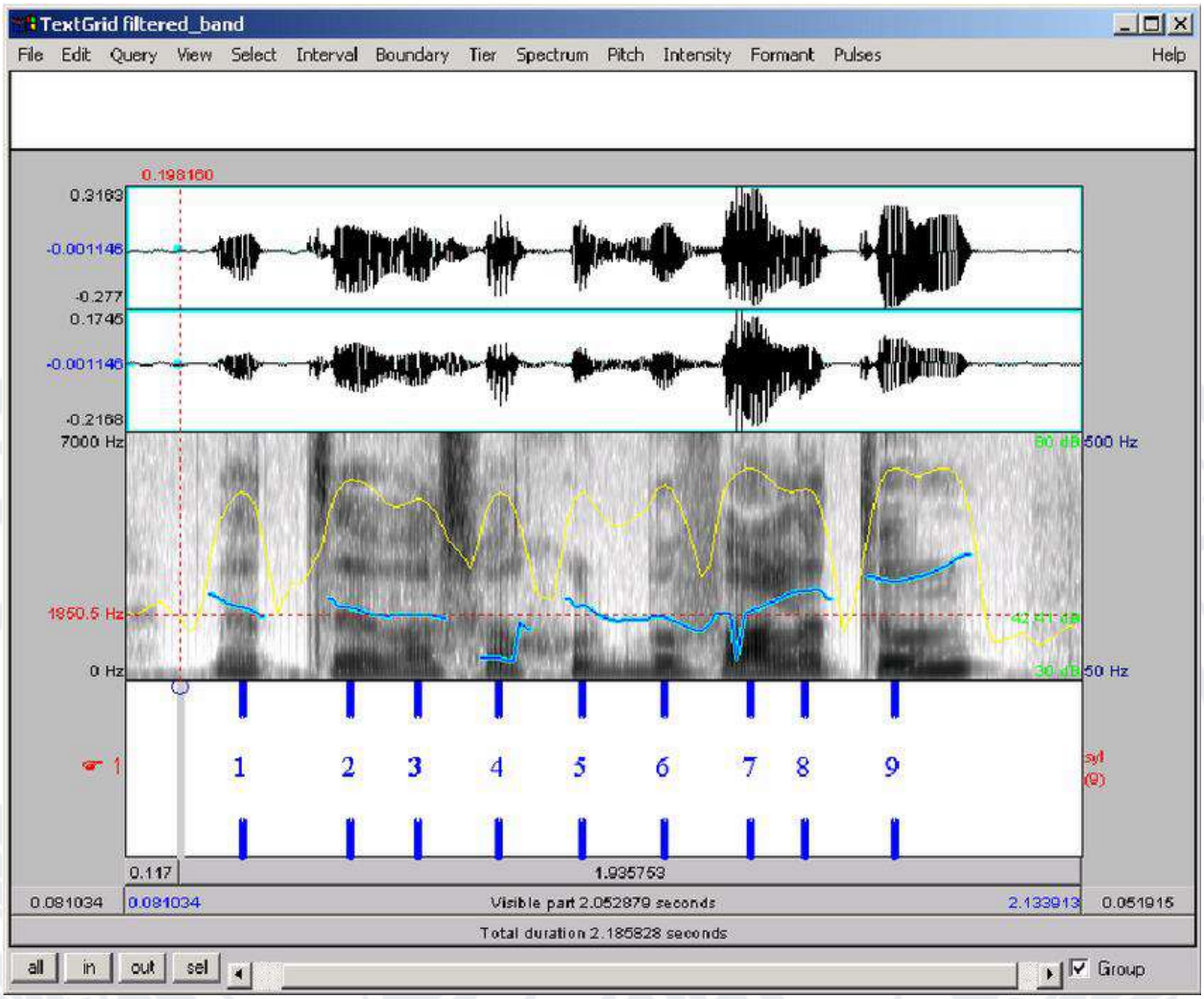


Рисунок 1.2 – Формат виводу результатів у Praat

Wavesurfer - це безкоштовне програмне забезпечення для аналізу звуку та сигналів. Воно надає користувачеві можливість візуалізації, аналізу та маніпуляції звукових даних [8].

Ось кілька основних особливостей Wavesurfer:

– Візуалізація звукових файлів: Wavesurfer відображає звуковий сигнал у вигляді графіка, що дозволяє користувачам легко візуально аналізувати форму хвилі та структуру звукового файлу;

– Аналіз звуку: Wavesurfer пропонує набір інструментів для аналізу звуку, включаючи спектральний аналіз, спектрограму, формантний аналіз та багато іншого. Ці функції дозволяють вивчати спектральні характеристики звуку та його акустичні властивості;

– Маніпуляція звуковими даними: Wavesurfer дозволяє користувачам виконувати різні операції зі звуковими файлами, такі як обрізка, наложення ефектів, зміна швидкості відтворення тощо. Це корисно при підготовці звукових даних для аналізу;
– Спектральний аналіз: Wavesurfer дозволяє виконувати спектральний аналіз звуку, що дозволяє визначити частотний склад звукового сигналу. Ви можете аналізувати спектрограми, форманти, шумові профілі та інші спектральні характеристики;
– Аудіозапис та відтворення: Wavesurfer дозволяє записувати звукові дані безпосередньо з мікрофону або імпортувати існуючі аудіофайли. Ви також можете відтворювати звук у реальному часі або в зміненому темпі;
– Розширюваність: Wavesurfer пропонує розширювану архітектуру плагінів, що дозволяє користувачам додавати додаткові функції та алгоритми аналізу звуку. Це дозволяє налаштовувати Wavesurfer під свої індивідуальні потреби;
– Кросплатформовість: Wavesurfer доступний на різних операційних системах, включаючи Windows, macOS та Linux, що дає можливість використовувати його на будь-якому пристрої;
– Підтримка різних форматів: Wavesurfer підтримує багато поширених аудіоформатів, таких як WAV, MP3, Ogg та інші, що робить його зручним для роботи зі звуковими файлами різного походження.

Wavesurfer є потужним інструментом для аналізу звуку та сигналів, який може бути корисним у різних галузях, включаючи лінгвістику, фонетику, акустику, звукозапис та інші. Завдяки своїй гнучкості та розширюваності, Wavesurfer надає користувачеві широкі можливості для роботи зі звуковими даними (рисунок 1.3).

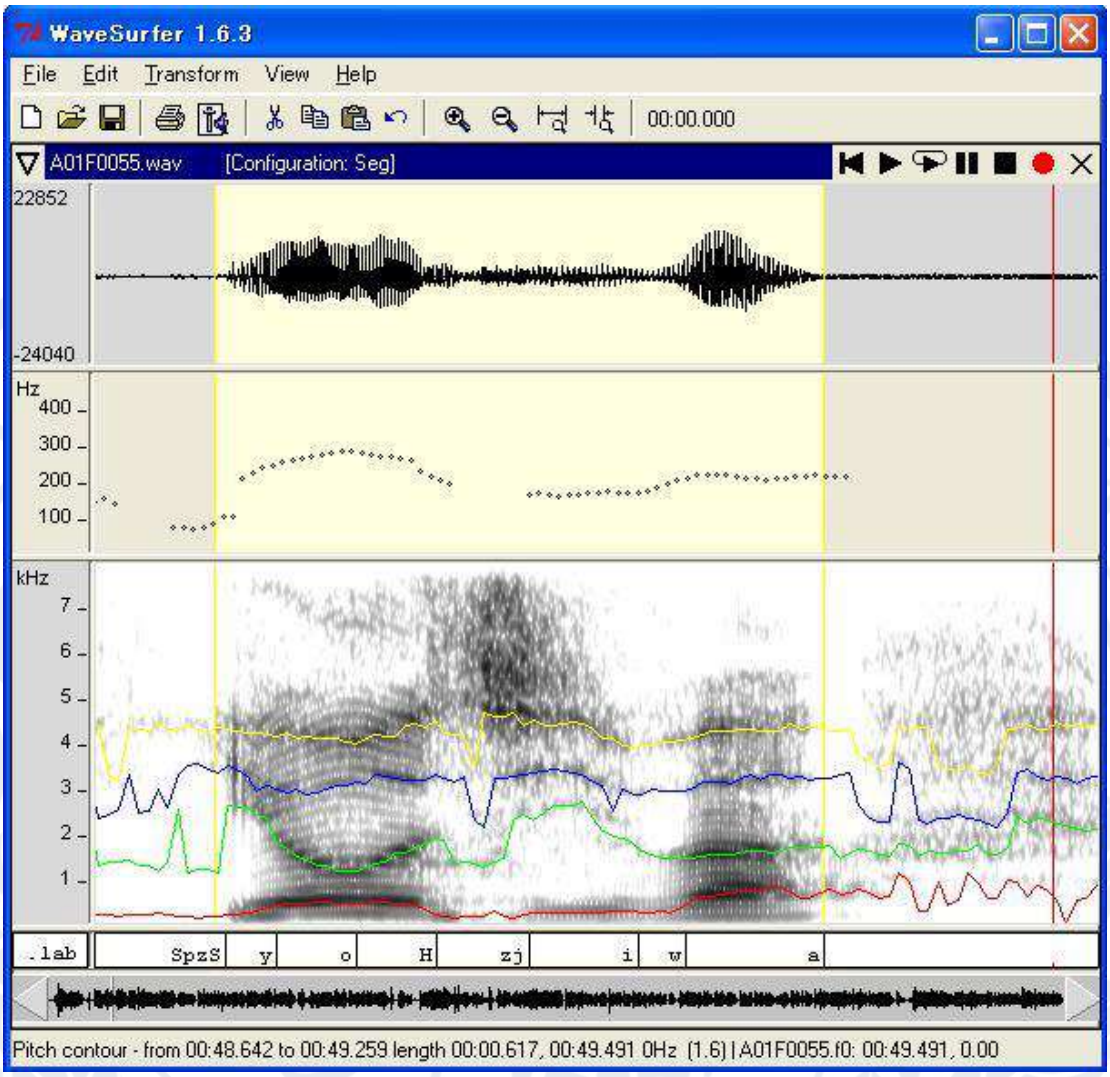


Рисунок 1.3 – Інтерфейс Wavesurfer

Переваги Wavesurfer:

1. Безкоштовність: Wavesurfer є безкоштовним програмним забезпеченням, що робить його доступним для широкого кола користувачів без необхідності витрат на ліцензію;
2. Візуалізація звуку: Wavesurfer надає детальну візуалізацію звукових файлів у вигляді графіків, що дозволяє користувачам швидко і зручно аналізувати форму хвилі та інші характеристики звукових сигналів;
3. Гнучкість і розширюваність: Wavesurfer має модульну архітектуру і підтримку плагінів, що дозволяє користувачам налаштовувати і розширювати функціонал програми відповідно до своїх потреб;

4. Підтримка різних форматів: Wavesurfer підтримує широкий спектр аудіоформатів, що дозволяє працювати зі звуковими файлами різних типів без необхідності використання додаткових конвертерів або кодеків.

Недоліки Wavesurfer:

1. Інтерфейс користувача: Деяким користувачам може знадобитись деякий час для ознайомлення з інтерфейсом Wavesurfer, оскільки він може виглядати складним або неінтуїтивним на початковому етапі;
2. Обмежений функціонал: У порівнянні з іншими аналогами, Wavesurfer може мати меншу кількість вбудованих функцій і інструментів для аналізу звуку. Це може бути обмеженням для деяких користувачів з більш специфічними потребами;
3. Вимоги до обладнання: Wavesurfer може вимагати певних ресурсів обчислювальної системи, особливо при роботі з великими або складними звуковими файлами. Недостатньо потужний комп'ютер може вплинути на продуктивність програми [9].

Варто враховувати, що переваги та недоліки Wavesurfer можуть бути індивідуальними і залежати від потреб і вимог кожного користувача.

1.4 Аналіз методів розв’язання поставленої задачі

Найшвидшим методом вирішення поставленої задачі буде модифікувати одну з уже існуючих систем, та додавання до неї нового функціоналу. Але, у такому разі, більша частина роботи покладеться саме на розробників вже існуючого програмного забезпечення. В іншому випадку, якщо завдання модифікації покладається на інших осіб, виникає ряд проблем, що може ускладнити цей процес.

Модифікація існуючого програмного забезпечення може бути складним завданням залежно від різних факторів. Ось декілька можливих складнощів, з якими можна зіткнутися під час модифікації програмного забезпечення:

1. Складність коду: Якщо код існуючої програми є заплутаним, погано структурованим або погано задокументованим, то модифікація може бути складним процесом. Важко зрозуміти, які частини коду впливають на певну функціональність і які зміни можна внести, не ризикуючи порушити інші частини програми;
2. Залежності: Існуюче програмне забезпечення може мати залежності від інших компонентів, бібліотек або зовнішніх сервісів. Модифікація програми може вимагати змін у цих залежностях, що може призвести до проблем зі сумісністю або конфліктів версій;
3. Недостатня документація: Відсутність або недостатня документація про функціональність, структуру або архітектуру програмного забезпечення може ускладнити процес модифікації. Без відповідних вказівок може бути важко зрозуміти, як працює програма і які зміни потрібно внести;
4. Відсутність тестів: Якщо існуюче програмне забезпечення не має вичерпних тестів, то модифікація може бути ризикованою. Внесення змін може привести до появи нових помилок або порушити належне функціонування існуючих функцій. Тестування модифікованого програмного забезпечення є важливою частиною процесу, але відсутність наявних тестів може ускладнити цю задачу;
5. Обмеження архітектури: Іноді існуюча архітектура програмного забезпечення може бути обмеженою і не передбачати певних змін або додаткової функціональності. У такому випадку модифікація може потребувати переробки архітектури або різних компонентів програми, що може бути дуже складною задачею.

Тому краще обрати інший варіант виконання завдання [10].

Альтернативним варіантом є розробка власного програмного продукту для втілення усіх задумів. Цей підхід надає можливість приділяти більше уваги конкретним модулям додатку. Ще одна перевага такого підходу – автономність і незалежність від інших існуючих систем, що підвищує надійність роботи програмного продукту, знижуючи ризик виходу його з ладу [11].

У випадку розробки повноцінної власної системи зникає необхідність адаптувати додаток під різні потреби сторонніх систем. Даний підхід теж має свій недолік, оскільки потребує повноцінної розробки нової системи, а тому є більш трудомістким.

Розробка програмного забезпечення охоплює проблеми якості, вартості та надійності. Деякі програми містять мільйони рядків коду, які, як очікується, повинні правильно виконуватися в умовах, що динамічно змінюються. Також кожна наступна модифікація вже існуючого ПЗ буде зі все більшим шансом буде тягнути за собою виникнення багів, та різних несумісностей коду. Через це складність створення ПЗ можна порівняти зі складністю конструювання складних сучасних машин, таких як літаки, потяги, автомобілі, тощо [12].

Врахувавши переваги та недоліки кожного методу було вирішено скористатися методом розробки власного програмного. Це дозволить створити повноцінний програмний продукт без необхідності залежності додатку від сторонніх систем.

Висновок до розділу 1

У даному розділі було розглянуто актуальність проблеми аналізу аудіосигналу, постановку задачі роботи, а також проведено огляд існуючих програмних продуктів даної тематики, були окреслені їх переваги та недоліки. Наступний розділ буде присвячений аналізу існуючих методів аналізу аудіоданих.

РОЗДІЛ 2
АНАЛІЗ ІСНУЮЧИХ МАТЕМАТИЧНИХ МОДЕЛЕЙ ТА МЕТОДІВ АНАЛІЗУ АУДІОСИГНАЛУ

2.1 Аналіз існуючих методів аналізу аудіосигналу

Аби розібратися з тим, який саме аналіз аудіосигналу потрібно провести, аби дізнатися, чи присутній на записі синтезований голос, потрібно для початку розглянути методи синтезу, та відштовхуватися від них.

Генерація людського голосу за допомогою штучного інтелекту включає в себе використання технологій аналізу природної мови (Natural Language Processing, NLP) та синтезу мови (Speech Synthesis) на основі даних, що були отримані в процесі аналізу. Цей процес відбувається наступним чином:

1. Збір даних: Збір аудіозаписів людей з різними голосовими характеристиками. Збір текстових даних для тренування моделей аналізу та синтезу природної мови;
2. Використання нейронних мереж для розуміння та аналізу природної мови, такої як розпізнавання слів, синтаксичний аналіз тощо;
3. Тренування моделі синтезу мови: Використання генеративних моделей для генерації аудіосигналів, які відповідають текстовій інформації;
4. Оптимізація якості голосу: Використання методів оптимізації та звучання голосу, щоб забезпечити природність та реалізм.

У відкритому доступі є декілька відкритих бібліотек та, які можна використовувати для створення систем синтезу мови. Фреймворків. Наприклад: Tacotron, WaveNet, та DeepVoice. Найефективнішою з-поміж них є нейронна мережа WaveNet, що дозволяє отримати високу якість голосу, який складно відрізнити від натурального. Але вона також найбільше навантажує апаратну частину. Щоб мати змогу працювати з нею потрібно мати потужнішу апаратну частину.

Натомість, для розпізнавання синтетично створеного голосу теж існують певні методи. Сучасні методи аналізу аудіосигналу використовуються для

розпізнавання та витягування характеристик звукових сигналів. Ось кілька ключових методів, які широко використовуються:

1. Short-Time Fourier Transform (STFT): Цей метод використовується для перетворення звукового сигналу з часової області в частотну. STFT розбиває сигнал на короткі часові вікна і виконує перетворення Фур'є для кожного вікна. Це дозволяє отримати інформацію про частотний склад сигналу з плином часу;
2. Mel-Frequency Cepstral Coefficients (MFCC): Цей метод використовується для витягування характеристик з голосових сигналів, які відображають спектральну інформацію звуку. Він моделює спектральні властивості людського слуху шляхом застосування фільтрації на основі Мел-шкали та використання перетворення кепстральних коефіцієнтів [13];
3. Wavelet Transform: Wavelet Transform використовується для аналізу аудіосигналів з точки зору часу та частоти. Вона використовує вейвлет-функції, які мають короткі часові та частотні властивості, що дозволяє отримати деталізовану інформацію про сигнал [14];
4. Deep Learning: Застосування глибокого навчання до аналізу аудіосигналів стає все популярнішим. За допомогою нейромережевих архітектур, таких як рекурентні нейронні мережі (RNN) або сверточні нейронні мережі (CNN), можна виконати задачі, такі як класифікація звуків, розпізнавання мови, розпізнавання музичних жанрів та інші.

2.2 Універсальна модель загального аналізу аудіосигналу

Сигнали, що надходять від об'єктів реального світу, можуть змінюватись в залежності від об'єктивних обставин зовнішнього світу та фізичного стану цих об'єктів. Системи розпізнавання сигналів, що ґрунтуються на жорстких алгоритмах, характеризується високими ймовірностями помилки через присутність фонового шуму. Фоновий шум завжди присутній в навколишньому світі. Якщо навіть у повністю звукоізольовану кімнату поставити мікрофон, та зробити запис так званої «тиші», то прослуховуючи запис все одно можна помітити певні шуми.

Причинами цього будуть мікро спотворення сигналу кабелем, що сприймає радіохвилі навколишнього середовища, а також рух повітря у кімнаті, що ні при яких обставинах не буває статичним.

Завжди найкращим рішенням буде зводити до мінімуму стороннні шуми ще на етапі запису аудіосигналу. Але, якщо шум все ж потрапив на запис, існують і програмні способи обробки звуку для зменшення його рівня до прийнятного для систем розпізнавання сигналів.

Структура системи розпізнавання сигналів (рис. 2.1) передбачає наявність трьох блоків, які можуть функціонувати в режимі розпізнавання або в режимі навчання, і охоплені зворотними зв'язками:

- Ідентифікація параметрів;
- Формування еталонів сигналів;
- Розпізнавання (порівняння з еталоном).

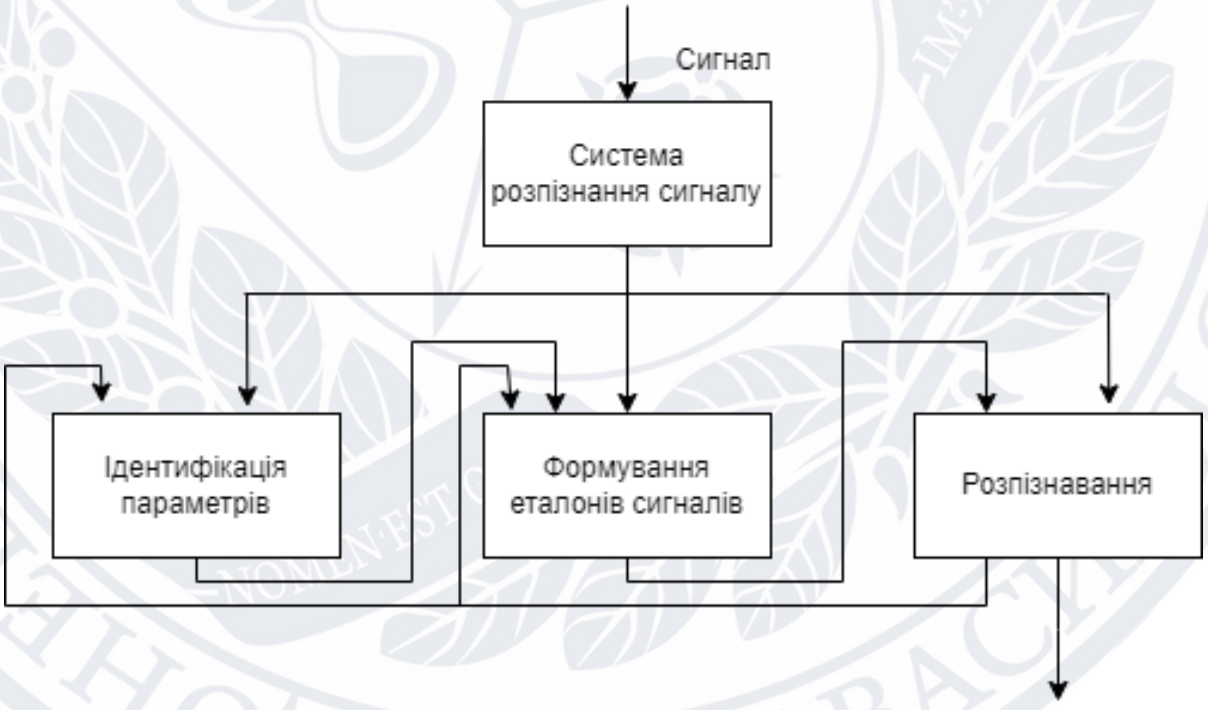


Рисунок 2.1 – Функціональна схема системи розпізнання

У першому блоці здійснюється ідентифікація параметрів системи. Цей режим визначає умови підготовки та налаштування параметрів, що використовуються при

формуванні еталонів та зіставленні з ними. Функціональна підготовка системи визначається у блоці формування еталонів. Цей блок призначений для персоніфікації програмного забезпечення системи згідно з особливостями сигналу. Реалізація сукупності перших двох блоків дозволяє функціонувати блоку зіставлення з стандартом, тобто. визначає ступінь функціональної готовності параметрів та бази знань для вирішення задач розпізнавання сигналів. У цьому блоці задаються умови ймовірності правильного зіставлення з стандартом, у разі невиконання яких система виходить із режиму зіставлення і вимагає додаткового навчання, тобто. перемикається в режим функціонування блоку формування стандартів сигналів. Блок ідентифікації параметрів вмикається при припиненні роботи двох інших блоків у тих випадках, коли проводиться зміна об'єкта, що досліджується, або заміна комплексу технічних засобів і стандартного програмного забезпечення системи.

Алгоритми розпізнавання сигналу створюються з урахуванням стандартів. Це зумовлює доцільність створення стандартів, що відбивають характерні особливості сигналу.

На рисунку 2.2 зображено узагальнену структуру блоку формування еталонів системи розпізнавання сигналів.

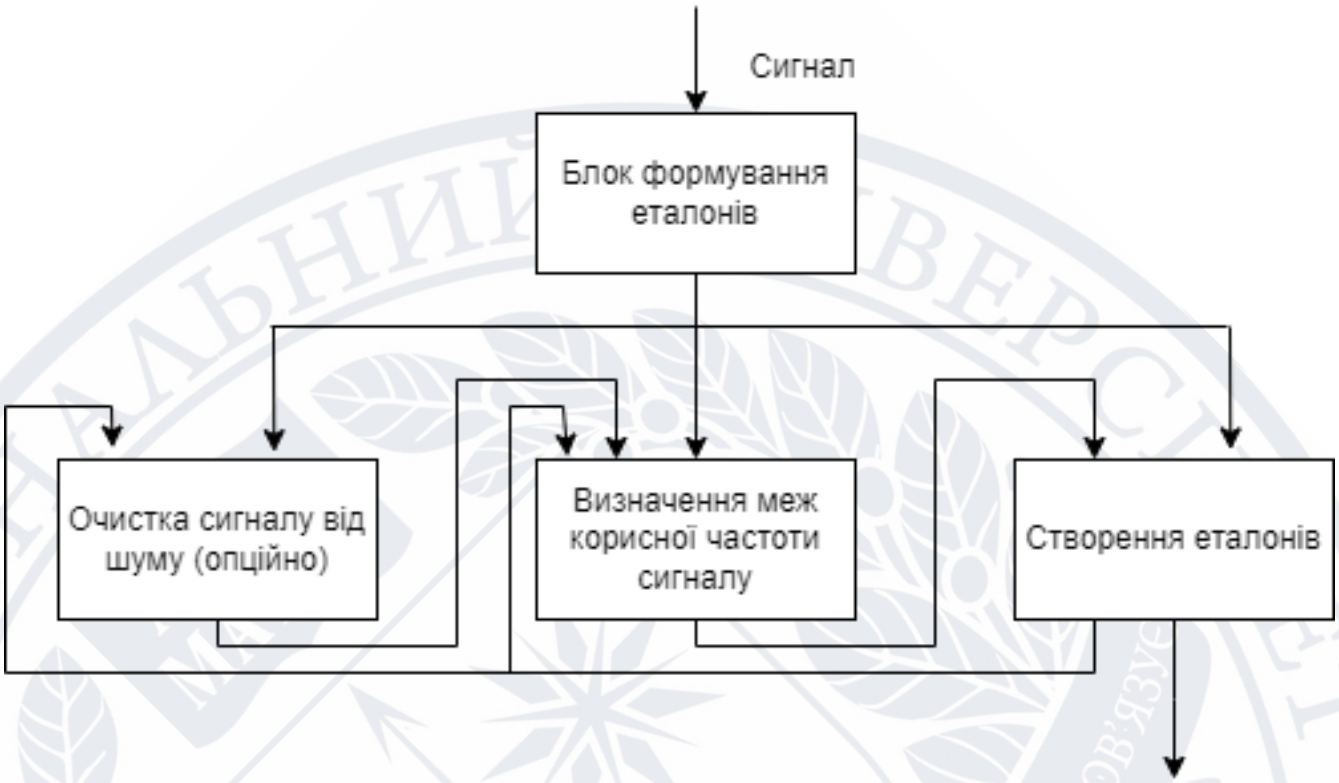


Рисунок 2.2 – Загальна структура блоку формування еталонів

На першому етапі використовуються методичні правила очищення сигналу від шуму, що враховують чисельні характеристики сигналу.

На другому етапі використовуються методичні правила визначення меж корисної частини сигналу, враховують чисельні характеристики сигналу.

На третьому етапі використовуються методичні правила створення бази даних еталонів сигналів, характеристики яких виділяються методами цифрової обробки.

Алгоритми розпізнавання сигналу базуються на зіставленні сигналу з еталонами, створеними попереднім блоком.

На рисунку 2.3 зображено узагальнену структуру блоку розпізнавання сигналів.

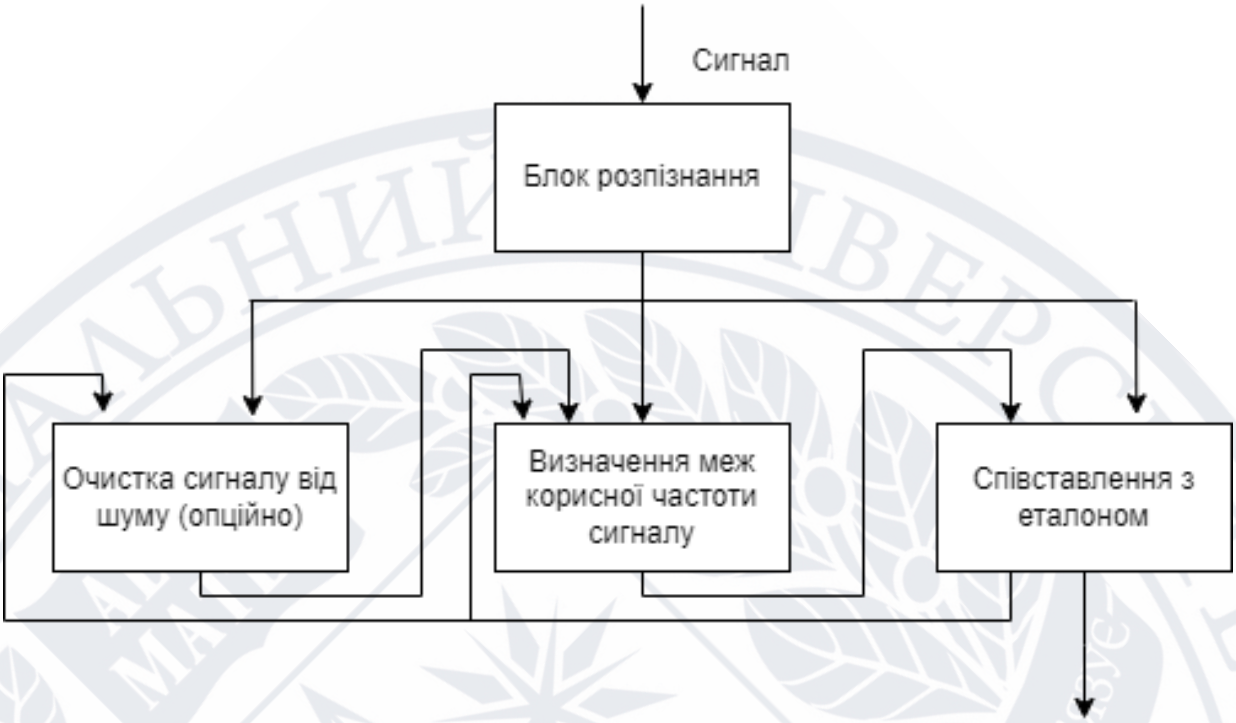


Рисунок 2.3 – Загальна структура блоку розпізнання

На першому етапі використовуються методичні правила очищення сигналу від шуму, враховують чисельні характеристики сигналу.

На другому етапі використовуються методичні правила визначення меж корисної частини сигналу, враховують чисельні характеристики сигналу.

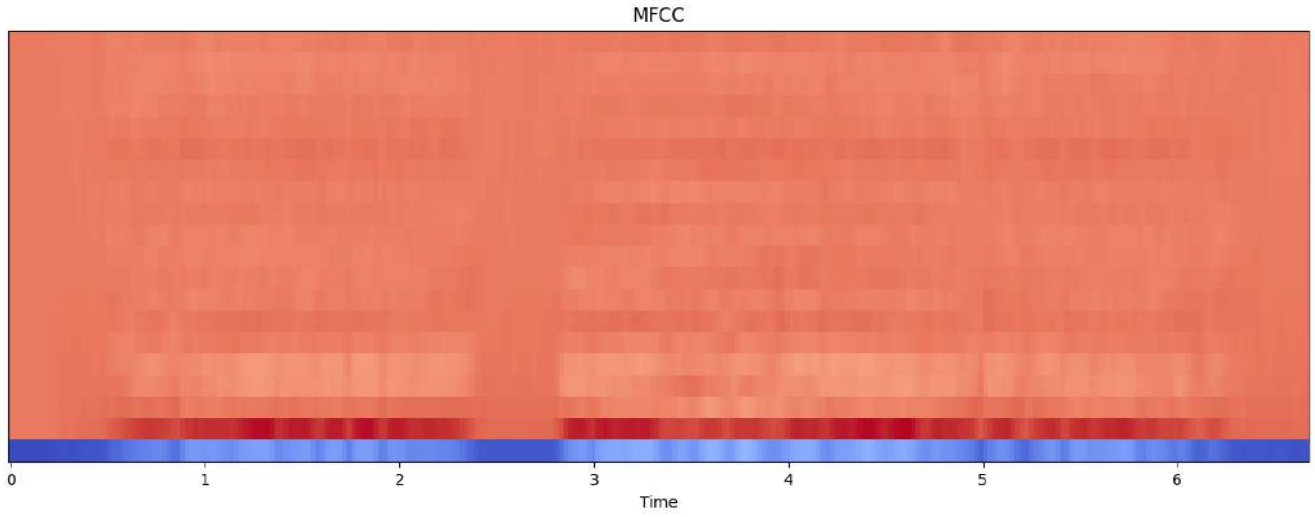
На третьому етапі використовуються методичні правила зіставлення сигналу, що розпізнається, з еталоном, характеристики яких виділяються методами цифрової обробки [15].

Відповідно до структури систем розпізнавання сигналів необхідно розглянути методичні положення, на яких базуються блоки формування еталонів та розпізнавання цих систем.

Найчастіше для ідентифікації синтетичного голосу використовується метод, що базується на визначенні Мел-кепстральних коефіцієнтів (MFCC). Саме він лежатиме в основі розроблюваного алгоритму [16].

Мел-шкала, це емпірична шкала, що ґрунтується на людському відчутті частоти звуку, була запропонована Стівенсом і Волкман в 1937 р [17]. Шкала була отримана в результаті експериментів, в яких, випробовуваних просили скорегувати

сигнал, який вони чують таким чином, щоб його висота стала в 2 рази нижчою. В результаті була отримана шкала, в якій 1000 Мел відповідає «висоті» звуку з частотою 1 кГц і подвоєння Мел створює відчуття сприйняття подвоєння висоти звуку за рахунок утворення обертонів. На рисунку 2.4 можна побачити вигляд діаграми Мел-кепстральних коефіцієнтів, яку можна побудувати за допомогою бібліотеки Librosa для Python.



Для обрахунку Мел-кепстральних коефіцієнтів для початку потрібно розкласти аудіосигнал на звукові хвилі (рисунок 2.5), а їх у свою чергу на перкусійні та гармонічні (рисунок 2.6). Після цього в дію вступає алгоритм обчислення.

У цьому методі як ознаки використовуються мелчастотні кепстральні коефіцієнти

1. Сигнал $s(m)$ розбивається на L фреймів завдовжки ΔN. Для n-го кадру виконується балансування спектра, що має крутий спад в області високих частот. $\breve{s}_n(m) = s_n(m+1) - \alpha s_n(m),\ m \in \overline{0, \Delta N - 1}$, де α – параметр фільтрації, $0 < \alpha < 1$.

2. Для n-го кадру обчислюється спектр: $\widehat{s}_n(m) = \breve{s}_n(m)w(m)$,

$$w(m) = 0.54 + 0.46\cos\frac{2\pi m}{\Delta N},$$

$$\widehat{S}_n(k) = \sum_{m=0}^{\Delta N-1} \widehat{s}_n(m) e^{-j(2\pi/\Delta N)km}, \quad k \in \overline{0, \Delta N - 1},$$

де $w(m)$ - вікно Хеммінга

3. Для n-го кадру на і-й мелчастотній смузі обчислюється енергія мелчастотних смуг, використовуючи перетворення частот і вікно Бартлета

$$\widehat{E}_{nm} = \sum_{k=0}^{\Delta N/2-1} \left|\widehat{S}_n(k)\right|^2 w_m(k), \; m \in \overline{1, P},$$

$$w_m(k) = \begin{cases} 0, & k < f_{m-1} \vee k > f_{m+1} \\ \dfrac{k - f_{m-1}}{f_m - f_{m-1}}, & f_{m-1} \le k \le f_m \\ \dfrac{f_{m+1} - k}{f_{m+1} - f_m}, & f_m \le k \le f_{m+1} \end{cases},$$

$$f_m = \frac{N}{f_d} B^{-1}\left(B(f^{\min}) + m\frac{B(f^{\max}) - B(f^{\min})}{P+1} \right),$$

$m \in \overline{0, P+1}$,

$B(f) = 1125\ln(1 + f/700)$,

$B^{-1}(b) = 700(\exp(b/1125) - 1)$,

де $\widehat{E}_{im}$ - енергія m-ї мелчастотної смуги

$w_m(k)$ - вікно Бартлета для m-ї полоси,

$B(f)$ - функція, яка перетворює частоту в Гц на мелчастоту,

$B^{-1}(b)$ - функція, яка перетворює мелчастоту на частоту в Гц,

f_m - нормувальна частота,

$f^{\min}$, $f^{\max}$ - мінімальна та максимальна частота в Гц,

f_d - частота дискретизації звукового сигналу в Гц,

P – кількість мелчастотних смуг.

4. Для n-го кадру обчислюються мелчастотні кепстральні коефіцієнти, використовуючи зворотне дискретне косинусне перетворення типу DCT-2.

$$MFCC_n(m) = \sqrt{\frac{2}{P}} \sum_{k=0}^{P-1} \ln\left(\widehat{E}_{n,m+1}\right) \alpha(k) \cos\left(\frac{(2m+1)k\pi}{2P}\right),$$

$m \in \overline{0, \widetilde{P}-1}$, $\alpha(k) = \begin{cases} \sqrt{\frac{1}{2}}, & k = 0 \\ 1, & k > 0 \end{cases}$, де $\widetilde{P}$ - кількість мелчастотних кепстральних коефіцієнтів, $1 \le \widetilde{P} \le P$.

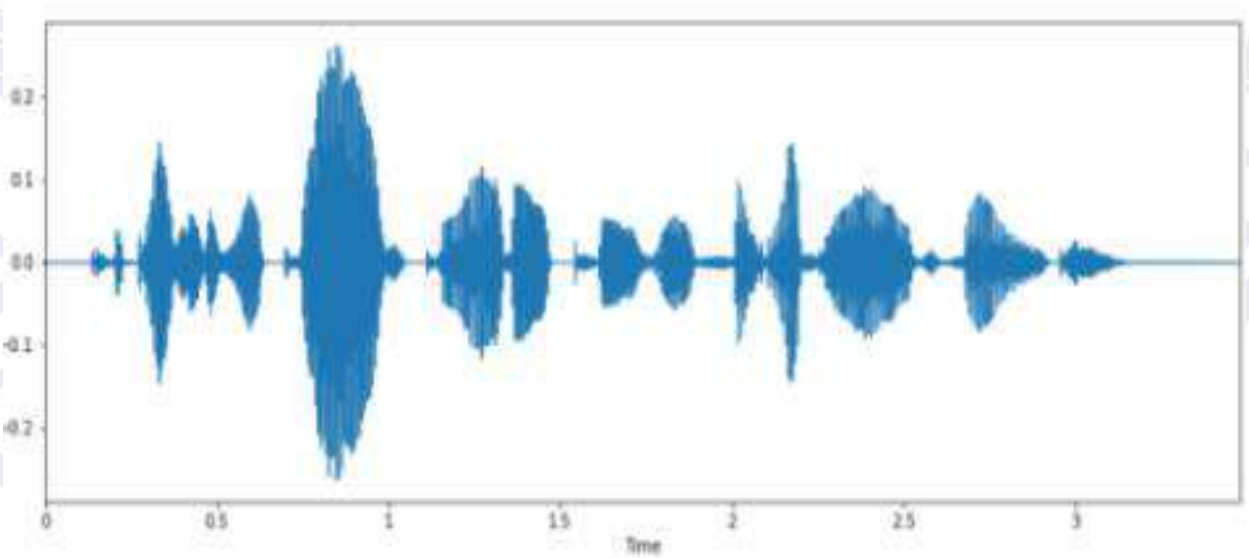


Рисунок 2.5 – Візуалізація розкладення аудіосигналу на звукові хвилі

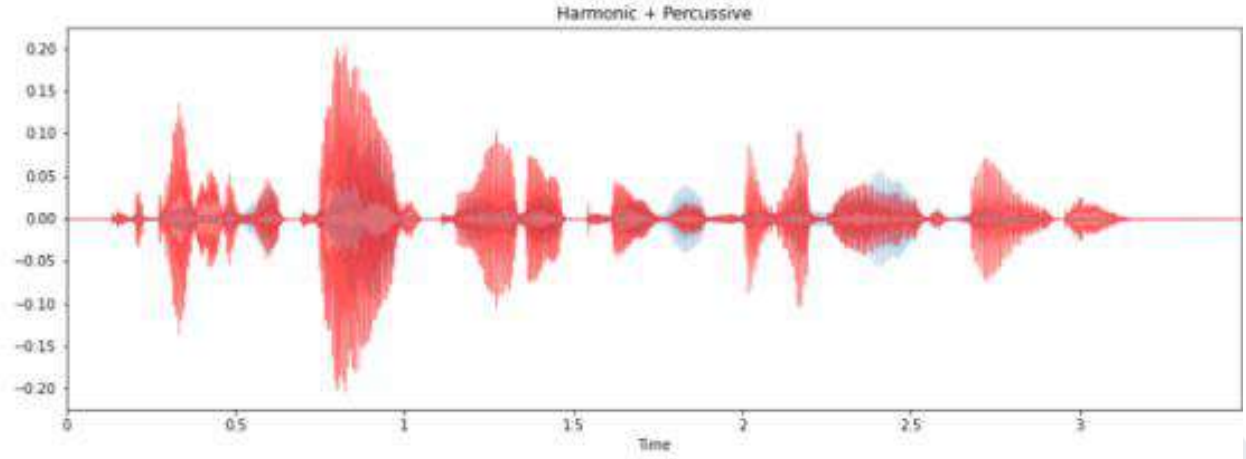


Рисунок 2.6 – Візуалізація розкладення звукових хвиль на гармонічні та перкусійні

За умови навчання і тестування без наявності шумів такий метод показав гарні результати точності визначення (94,31 % – 96,32 %), але точність розпізнавання значно погіршується з введенням шумів у тестовий аудіозапис (56,26 % – 74,27 % при відношенні сигнал/шум 10 дБ). Існують і альтернативні алгоритми та методи аналізу аудіосигналу на аналогічні елементи. Найближчим алгоритмом до використання Мел-кепстральних коефіцієнтів (MFCC) є використання Барк-кепстральних коефіцієнтів (BFCC). Для рівномірної Мел-кепстральної шкали найбільший програш у точності розпізнавання досяг 8.27 % відносно точності найближчої шкали (BFCC, шум за експерименту – виставковий зал, відношення сигнал/шум 10 дБ). При відсутності ж фонового шуму, перевагу має використання методу аналізу Мел-кепстральних коефіцієнтів [18].

2.3 Дослідження змішаного типу генерації штучного людського голосу та аналіз можливостей його ідентифікації на записі

Також для ідентифікації синтетичного голосу дуже часто є використання спектрограм.

Спектрограма – це візуальне зображення спектру частот сигналу в часі. Спектрограми використовуються для ідентифікації та обробки речі, аналізу звуків тварин, у різноманітних сферах музики, радіо- та гідролокації, сейсмології та інших

областях. При застосуванні до звукового сигналу спектрограми іноді називають сонографами, голосовими відбитками або голосограмами (рисунок 2.7) [19]. Коли дані представлені в тривимірному графіку, вони можуть називатися водоспадами.

Спектрограми часто використовуються для аналізу звучання музичного інструменту. Так можна легко виявити його переваги та недоліки, а також можливі похибки у конструкції. Також спектрограми можна використовувати для аналізу голосу людини. Відбиток голосу виступає найбільш унікальною формою розпізнавання людини серед інших. Така форма більш складна і неповторна, ніж відбиток пальця чи сітківка ока. Відбиток пальця являє собою лише точне відтворення ліній шкіри, в той час як відбиток голосу – це поєднання вимовного акценту, специфіки вимови наголошених та ненаголошених складів, ритміки висловлювання, закінчень слів або фраз та інше, що пов`язано з формою, силою та розміром голосових зв`язок мовця [20].

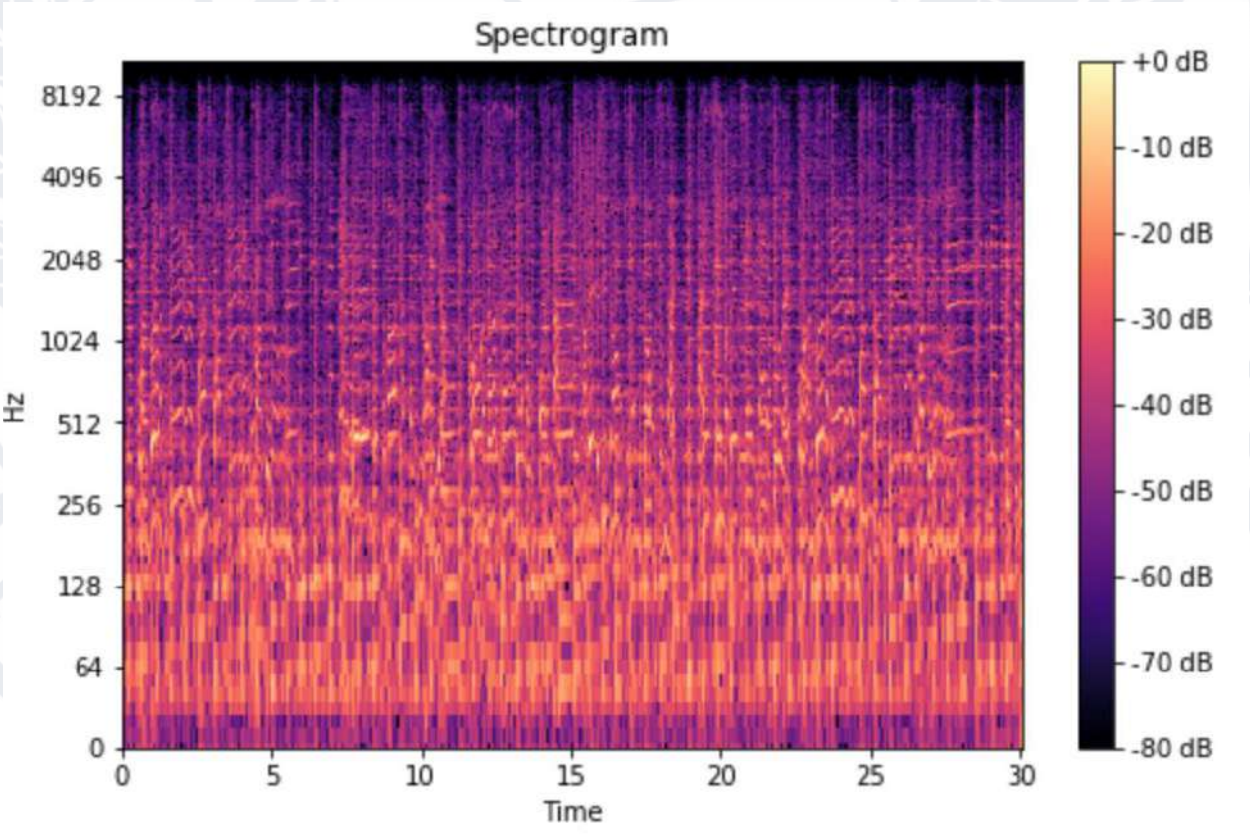


Рисунок 2.7 – Двовимірна спектрограма

Найбільш поширеним уявленням спектрограми є двовимірна діаграма: на горизонтальній осі представлено час, вертикальної осі - частота; третій вимір із зазначенням амплітуди на певній частоті в конкретний момент часу представлено інтенсивністю або кольором кожної точки зображення. Частота та амплітуда осей може бути лінійними чи логарифмічними, залежно від того, з якою метою використовується графік. Аудіо зазвичай може бути представлено з логарифмічною віссю амплітуди (часто, в децибелах або дБ), і частота буде лінійною, щоб підкреслити гармонійні відносини, або логарифмічної, щоб підкреслити тональні характеристики аудіосигналу.

Зазвичай використовується один з двох способів створення спектрограми:

1. Сигнал апроксимується, як набір фільтрів, отриманих із серії смугових фільтрів (це був єдиний спосіб до появи сучасних методів цифрової обробки сигналів);
2. Розраховується за сигналом часу, використовуючи віконне або швидке перетворення Фур'є.

Ці два способи фактично утворюють різні квадратичні частотно-тимчасові розподіли, але еквівалентні за певних умов. Метод смугових фільтрів зазвичай використовують у аналогової обробці для поділу вхідного сигналу на частотні діапазони.

Створення спектрограм за допомогою віконного перетворення Фур'є зазвичай виконується методами цифрової обробки. Проводиться цифрова вибірка даних у часовій області. Сигнал розбивається на частини, які зазвичай перекриваються, і потім проводиться перетворення Фур'є, щоб розрахувати величину частотного спектра для кожної частини. Кожна частина відповідає вертикальній лінії на зображенні – значення амплітуди в залежності від частоти у кожний момент часу. Спектри або часові графіки розташовуються поруч на двовимірній діаграмі [21].

Отже, спектрограма є способом представлення у зручному вигляді частотних (тональних) характеристик аудіосигналу. Спектрограма може бути використана для аналізу аудіозапису на наявність синтетичного голосу за допомогою

порівняння обертонів та гармонік, що відповідають за наявність тембру голосу у людини.

Тембр голосу – це індивідуальне, не схоже ні на кого звучання голосу кожної людини окремо, яке визначається основним тоном та додатковими звуками – обертонами. Чим більше обертонів, тим яскравішим є голос людини. [22].

Обертон — це будь-яка резонансна частота, яка перевищує основну частоту звуку. Якщо частоти обертонів у ціле число разів більші від частоти основного тону, то їх називають гармонічними обертонами (гармоніками) [23]. Іншими словами, обертони — це всі висоти, вищі за найнижчу висоту в окремому аудіосигналі. Основним є найнижчий тон. У той час як основний звук зазвичай чутний найбільш чітко, обертони фактично присутні в будь-якій висоті, крім справжньої синусоїди [24]. Відносна гучність або амплітуда різних часток обертону є однією з ключових ідентифікаційних ознак тембру, або індивідуальної характеристики звуку. Генерування обертонів та гармонік може відбуватися за допомогою різних інструментів. Стосовно теми генерації голосу їх слід поділити на два типи:

1. Інструменти для прикрашання природнього голосу;
2. Інструменти для генерації голосу.

Перший тип інструментів: ті, що допомогають прикрасити голос живої людини, додаючи до її природніх обертонів додаткові, що будуються на вже існуючих. Саме наявність базових обертонів допомагає голосу звучати яскравіше, та досить природньо навіть для програмного аналізу. Такий метод використовується при редагуванні запису голосу, та надання йому більш чітких та виразних частотних контурів. Він застосовується у різних культурних сферах життя, таких як телебачення, радіо, музика, запис аудіокниг, тощо.

Другий тип інстурментів використовується для генерації синтетичного голосу з нуля. Такі інструменти все ще генерують необхідно і достатньо природний рівень обертонів для неозброєного вуха людини. Але без початкових природних гармонік майже неможливо досягнути рівня звуку, що здатний пройти програмну перевірку аудіосигналу.

Справжнім викликом для системи аналізу аудіосигналу на предмет наявності на записі синтетичного голосу може стати змішаний тип генерації, який буде працювати на базі редагування записів голосу живої людини. Принцип роботи цього складеного методу наступний:

1. Відбувається підбір потрібних аудіоелементів з бази даних (записані слова або фрази, що промовляє одна й та сама жива людина);
2. Підібрані аудіоелементи розташовуються в певній послідовності;
3. Приведення всіх ділянок аудіосигналу до потрібного рівню, що відповідатиме певному значенню рівня гучності;
4. Прибирання пік-фактору сигналу за допомогою компресії аудіосигналу;
5. Визначення тональності голосу диктора, визначення середніх значень частоти на парах кінець одного уривку/початок наступного уривку;
6. Внесення тональних виправленнь за допомогою технології pitch shift.

Усі записи живої людини, навіть зроблені при ідентичних умовах, не можуть мати однаковий рівень гучності запису.

Гучність аудіосигналу найчастіше вимірюють у децибелах, але ця величина не є актуальною, коли справа стосується саме людського сприйняття звукових хвиль. У такому разі потрібно застосовувати іншу одиницю вимірювання, люфс (англ Lufs). Ці одиниці виміру схожі між собою, але мають деякі характерні відмінності.

Децибели (dB):

1. Децибели використовуються для вимірювання абсолютної або відносної величини гучності або сили сигналу;
2. Це логарифмічна одиниця вимірювання і використовується для вираження співвідношення між двома величинами;
3. У контексті аудіо децибели можуть використовуватися для вимірювання рівня гучності звукового сигналу, амплітуди аудіосигналу, відношення сигнал-шум і т.д.

Люфс (LUFS):

1. Люфс є абсолютною одиницею вимірювання гучності аудіосигналу, яка враховує специфічні характеристики слухового сприйняття людини;
2. Використовується для оцінки загальної гучності аудіозапису на основі середньовагованого виміру енергії звукового сигналу;
3. За допомогою одиниць Люфс можна враховувати динаміку аудіо та коригувати вимірювання відповідно до того, як звук чується людині.

Звук, що є однаковим по рівню у децибелах може сприйматися людиною по-різному в залежності від його частотних характеристик. Графік гучності Флетчера-Мунсона (Fletcher–Munson equal loudness contours) дозволяє побачити рівень сприйняття сигналу людиною в залежності від його гучності та тональної характеристики (рисунок 2.8). Таким чином, звуковий сигнал з частотою 20 Гц. людина почує лише якщо його рівень гучності буде вищим за 70 дБ. А сигнали з частотою 3-5 кГц. людина може вловити навіть якщо їх рівень нижчий за звуковий тиск, що відповідає еквіваленту 0 дБ (2×10−5 Па). Саме тому доречним буде використання одиниць вимірювання Люфс для виміру гучності аудіосигналу на записі.

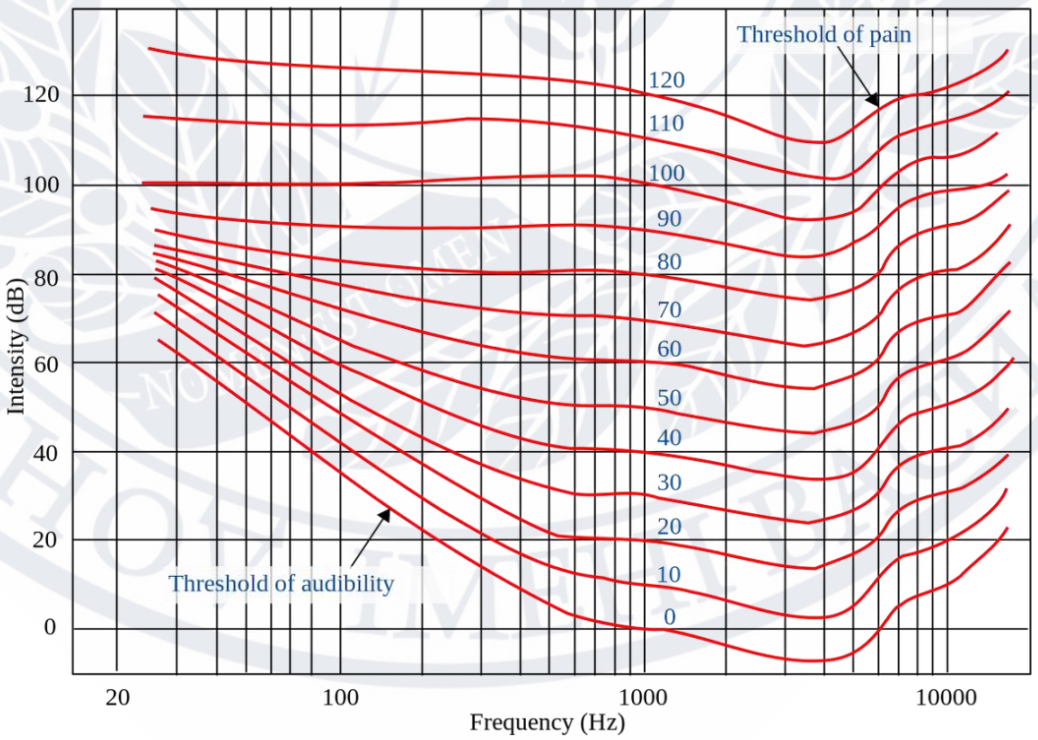


Рисунок 2.8 - Графік гучності Флетчера-Мунсона

Отже, першою нагальною проблемою стає привести всі відрізки запису до єдиного рівня. Цей процес виконується у 2 кроки. За допомогою аналізатора гучності збираються дані про гучність усіх аудіо відрізків з окремими словами або фразами. Прикладом аналізатора рівню гучності може бути youlean loudness meter (рисунок 2.9). Штучний інтелект користуватиметься скриптами та програмним кодом зі схожим принципом дії. Потім в одиницях вимірювання люфс задається потрібний рівень гучності запису, після чого рівень аудіосигналу підіймається, чи опускається. Для плавного поєднання кінців та початків відрізків використовується технологія fade in/fade out, що дозволяє плавно зменшувати або збільшувати гучність, та уникати різких стрибків її значення при переході на новий відрізок аудіо даних.

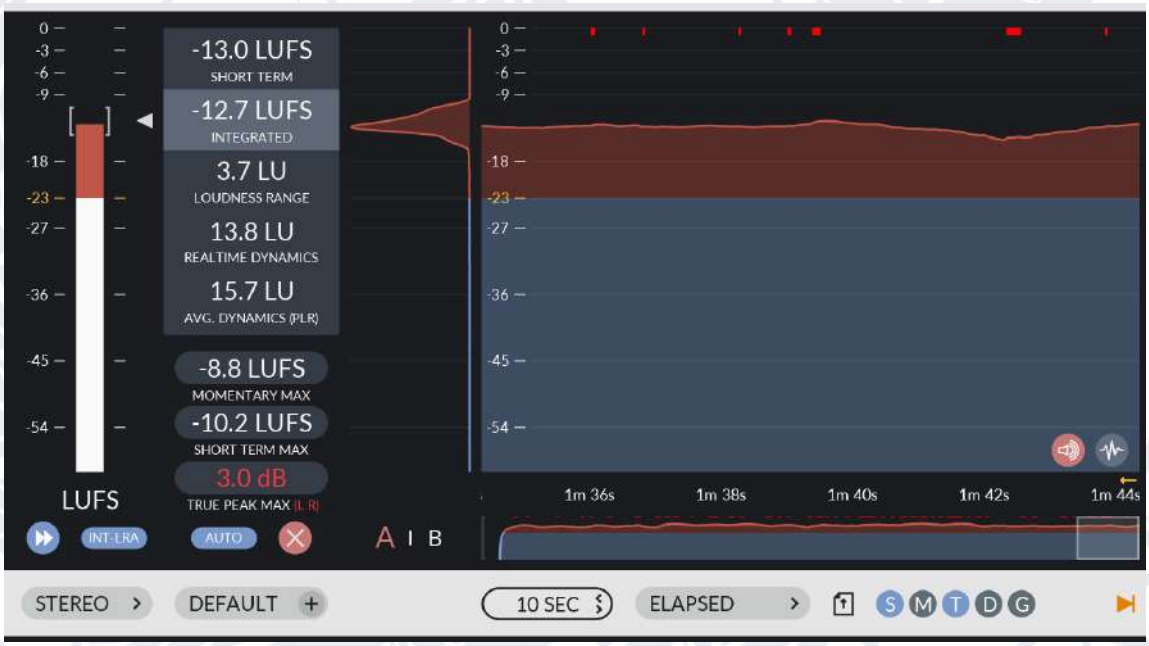


Рисунок 2.9 – Аналізатор рівню гучності “YOULEAN LOUDESS METER”

Далі потрібно позбавитися від пік-фактору запису. При збільшенні рівню гучності окремих відрізків звуку, часто на записі часто з`являються артефакти, від яких потрібно позбутися. За допомогою компресії аудіосигналу можна відсотково стиснути його, не зачепивши відрізки без артефактів (рисунок 2.10).

Компресія (стиснення) - процес зміни динаміки звуку, вирівнювання його гучності, що робить гучний звук тихіше. Компресор має 4 основних регулятори.

THRES, або Threshold, допомагає встановити рівень сигналу, що слугує тригером початок роботу компресора. Вище значення означає, що рівень сигналу має бути більшим для того щоб спрацював компресор і навпаки.

COMP або Compression (компресія) встановлює рівень компресії, який буде застосований до сигналу. Цей рівень разом із значенням параметра THRES визначає у скільки разів компресор стисне сигнал.

RESP або Response (віддача) визначає як компресор реагує на аудіосигнал. В крайньому правому положенні компресор працює як peak-limiter виконуючи прості аттенюації (зниження рівня сигналу), коли сигнал досягає встановленого ручкою THRES рівня сигналу.

Gain (посилення) – рівень підсилення звукового сигналу дозволяє компенсувати втрати в гучності.

Слід зауважити, що компресор стискає сигнал у залежно від того, наскільки вхідний сигнал перевищує заданий поріг. Якщо сигнал має рівень у -3 дБ, а поріг спрацювання компресора становить -6 дБ з коефіцієнтом стиснення 3, то вихідний аудіосигнал буде мати рівень -6 дБ + |(-6 дБ - (-3 дБ))| / 3 = -5 дБ (рисунок 2.11). Якщо для прибирання артефактів з запису потрібно встановити рівень сигналу до -6 дБ, або нижче, поріг спрацювання потрібно зсунути нижче. Якщо зсунути його до -9 дБ, то пікові моменти запису будуть мати гучність -9 дБ + |(-9 дБ - (-3 дБ))| / 3 = -7 дБ, що достатньо аби прибрати артефакти (рисунок 2.12).

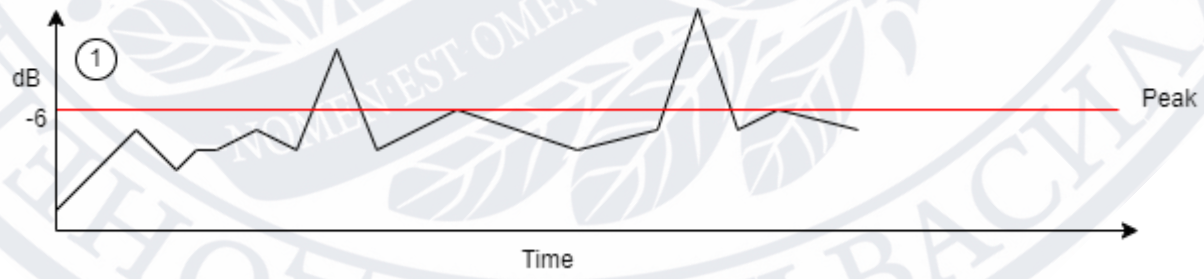


Рисунок 2.10 – Початковий рівень сигналу

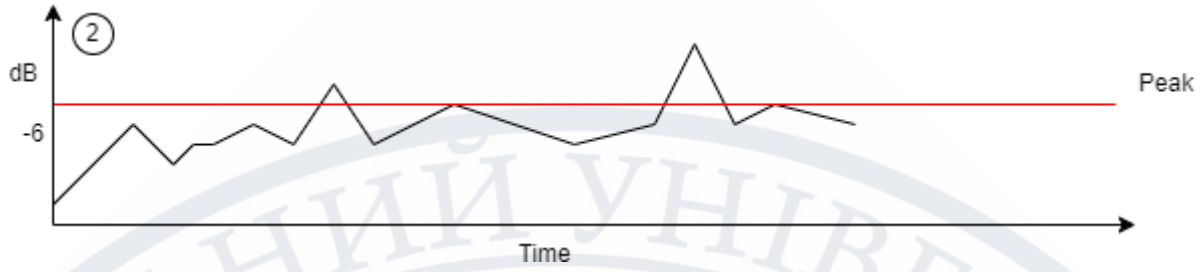


Рисунок 2.11 – Рівень сигналу після компресії (Threshold -6 dB)

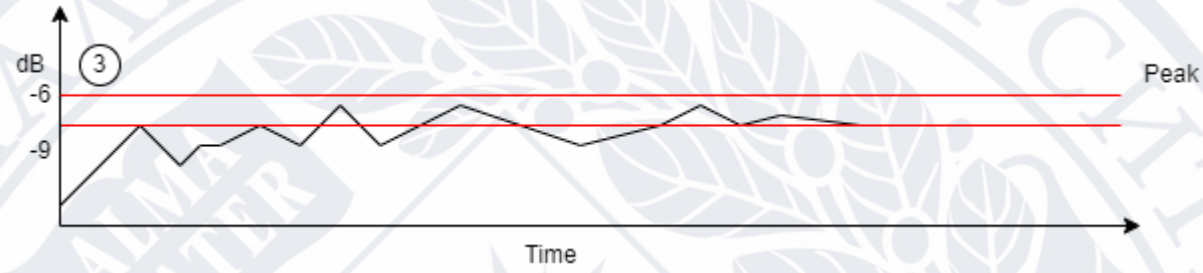


Рисунок 2.12 – Рівень сигналу після компресії (Threshold -9 dB)

Після того, як рівні гучності вирівняно та збалансовано, потрібно внести тональні виправлення аби різні відрізки звучали без частотних стрибків, як фраза, яка була суцільною початково [25]. Більшість програмного забезпечення типу pitch shifter має схожий принцип дії. А ще їх можна розділити ПЗ з можливістю редагування у реальному часі, та пз, що редагує вже записані фрагменти. Програмне забезпечення, що дозволяє вносити тональні виправлення в реальному часі, більш примітивне через потребу у швидкості його роботи. Саме тому воно підійде лише для того, щоб однаково підвищити, чи понизити тональність всього запису, та не є актуальним для задачі з конструювання синтетичного голосу змішаним методом, адже всі різкі тональні переходи не згладжуються. Тому актуальним для використання є інший тип, що відноситься до іншого типу. Їх алгоритм дій наступний:

1. Перш ніж отримати доступ до численних функцій тонального редагування, спочатку потрібно проаналізувати аудіоматеріал. Так як цей аналіз повинен торкнутися аудіофайлу цілком, його неможливо виконати у реальному часі. Аналіз виконується один раз відразу при передачі, після чого в нотному редакторі з'являться так звані «краплі». У випадку з плагіном все набагато складніше. Такому ПЗ потрібен всебічний аналіз,

тому необхідно заздалегідь передати сегменти аудіоданих, які потрібно відредагувати, щоб програма змогла їх проаналізувати. Цей процес називається "передача". Це процес запису, за допомогою чого ПЗ робить свою власну копію сегментів треку, які відтворюються у хості, що надав його. Таким чином програма отримує аудіодані, необхідні для аналізу та відображення нот (крапель). Процес передачі займає певний час та ресурси системи;

2. Після аналізу аудіоданих програмне забезпечення визначає основну тональність сегментів, а також середню тональність пар кінець одного голосового відрізку/початок іншого голосового відрізку;
3. Коли аналіз проведено, починається процес внесення тональних виправлень.

Навіть при такому методі залишається можливим виявити штучні виправлення та редагування голосу. Першочергово, саме через наявність на графіках рівнів гучності слідів компресії, що будуть виражатися як велика кількість досить пологих підвищень, та сходитимуться до одного й того ж рівня. Цього можна запобігти додавши голосу обертонів та гармонік, що зробить графік дещо більш схожим на такий, який відповідає природньому голосу. Однак, додаючи їх у діапазоні 3-5 тис. Гц. потрібно знову перевірити аудіосигнал за допомогою аналізатора гучності, оскільки згідно з графіком гучності Флетчера-Мунсона людське вухо дуже чутливе до сигналу такої частоти. А отже на записі можуть з`явитися нові чутні артефакти, які потрібно знову згладити.

В випадку, коли штучний голос згенеровано саме таким змішаним методом, потрібно звертатися до більш широкого аналізу спектру звуку, що надасть змогу виявити нехарактерні елементи для природного голосу людини.

Спектральний спад аудіосигналу - явище, що характеризує, як змінюються амплітуди різних частотних компонентів аудіосигналу з часом. У звуковому сигналі можуть бути присутні різні частоти, які представляють різні звукові компоненти. Спектральний спад описує те, як амплітуди цих компонентів зменшуються або спадають з плином часу. Такий аналіз може бути корисним при

дослідженні різних аспектів аудіосигналів, таких як характеристики звуку, динаміка звучання та інші. Він використовується у низці областей, та має свою специфіку використання у кожній із них:

1. У студійній роботі та музичній продукції спектральний спад може вказати на зміни в акустичних характеристиках записаного матеріалу. Це може допомогти виправляти або підсилювати певні частоти для досягнення бажаного звучання;
2. У області аудіоаналізу спектральний спад може використовуватися для визначення характеристик аудіозаписів, таких як тривалість відбуваючихся подій, зміни інтенсивності та інші аспекти;
3. У аудіокодуванні, такому як формати стиснення звуку (наприклад, при перетворенні початкового аудіоформату WAV на формат MP3), аналіз спектрального спаду може використовуватися для вилучення та збереження тільки найважливіших компонентів сигналу, що дозволяє ефективно стискати аудіодані, без надчуттєвих втрат для людського вуха;
4. Аудіологія: У медичній області, спектральний спад може бути використаний для дослідження слухових втрат та інших аудіологічних аспектів;
5. В звуковому дизайні, аналіз спектрального спаду може бути використаний для створення реалістичного та емоційно насиченого звучання голосу.

Один з варіантів використання спектрального спаду для аналізу аудіосигналу на предмет присутності на записі синтетичного голосу – дослідження динаміки запису, та пошук на ньому подібних патернів динаміки спектру. На рисунку 2.13 можна побачити діаграму спектрального спаду, яка побудована за допомогою бібліотеки Librosa для мови Python.

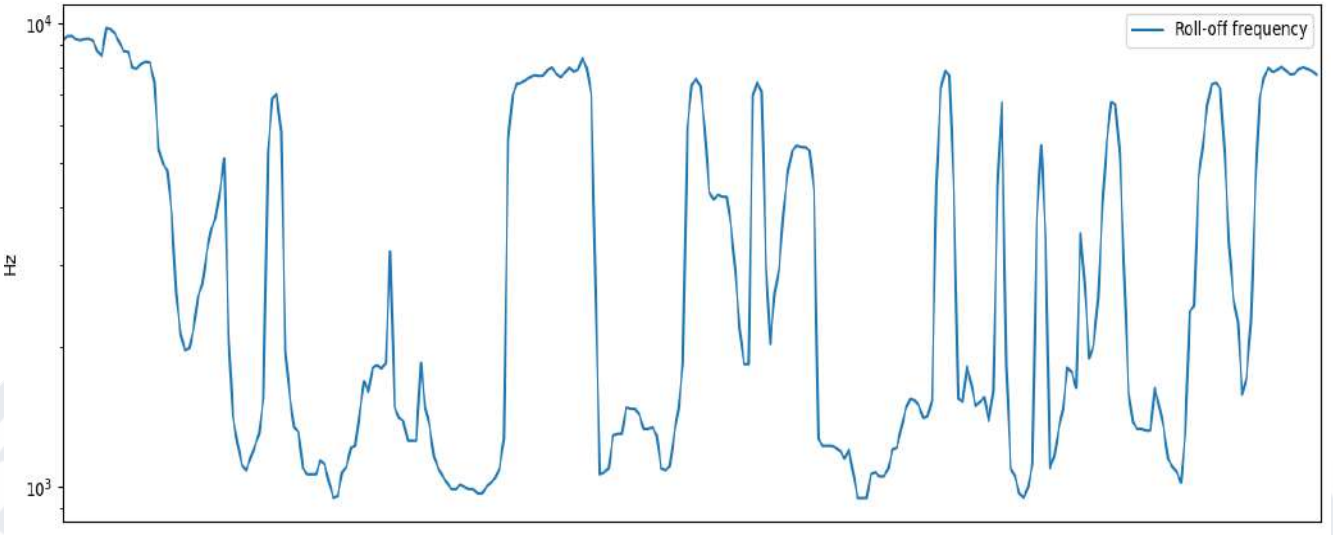


Рисунок 2.13 – Діаграма спектрального спаду

Спектральний центроїд (англ. Spectral Centroid) є одним з параметрів аудіосигналу, які визначаються для повного його спектрального аналізу. Це статистичний показник, який вказує на "центр мас" частотного спектру звукового сигналу та дозволяє отримати інформацію про те, де приблизно розташована "вага" спектральної енергії.

Математично спектральний центроїд визначається як середнє арифметичне частот, зважених їхніми амплітудами. Це можна обчислити за наступною формулою: $SC=\sum_k f_k A_k / \sum_k A_k$,

де:

- SC - спектральний центроїд;
- f_k - частота k-го біна (частотної складової) у спектрі;
- A_k - амплітуда k-го біна.

Спектральний центроїд може бути використаний для характеризації "звукового центру" аудіосигналу. Наприклад, якщо спектральний центроїд знаходиться високо, це може вказувати на те, що більше енергії зосереджено у високих частотах, що може сприяти враженню «яскравості» чи «шипіння» в аудіосигналі. З іншого боку, низький спектральний центроїд може вказувати на переважання низьких частот та присутності на аудіозаписі широкого басу або гудіння.

Спектральний центроїд має декілька конкретних застосувань у сфері обробки сигналів і аудіоаналізу:

1. Спектральний центроїд може служити показником того, де розташована основна "вага" частотної інформації в аудіосигналі. Високий спектральний центроїд може вказувати на велику кількість енергії в високих частотах, що може бути пов'язано з яскравим, гострим або шиплячим звучанням;
2. У музичній індустрії спектральний центроїд використовується для аналізу та характеризації аудіосигналів. Він допомагає в розпізнаванні музичних інструментів, а також оцінці тембральних особливостей та визначенні характеристик вокалу, що свідчить про те, що у сфері аналізу голосу він має пряме використання;
3. Аудіообробка і обробка сигналів: У сфері обробки сигналів спектральний центроїд може бути використаний для автоматизації аудіообробки, такої як еквалізація (регулювання рівнів частот) або компресія (керування динамікою);
4. В аудіорозпізнаванні спектральний центроїд може бути використаний для визначення характеристик мовлення. Різні мовленнєві звуки можуть мати різні спектральні центроїди, що робить його корисним параметром для розпізнавання мовлення для подальшого його перетворення на текст та для аналізу аудіосигналів мовлення;
5. У галузі медичної акустики спектральний центроїд може бути використаний для вивчення властивостей акустичних сигналів, пов'язаних з фізіологічними процесами організму людини, такими як серцебиття чи артеріальний тиск.

На рисунку 2.14 можна побачити діаграму спектрального центроїду, яка побудована за допомогою бібліотеки Librosa для мови Python.

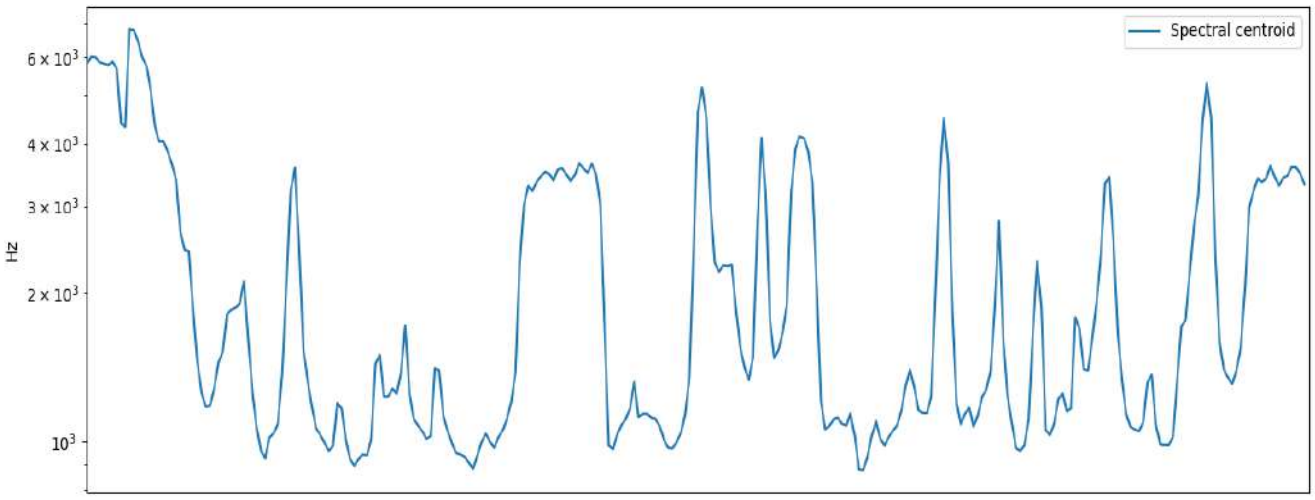


Рисунок 2.14 – Діаграма спектрального центроїду

Спектральний контраст аудіосигналу вказує на різницю у спектральній енергії між різними частотними діапазонами в аудіосигналі. Цей параметр використовується для характеризації того, наскільки сильно відрізняються амплітуди різних частот у звуковому сигналі.

Математично спектральний контраст може визначатися різними способами, але в основі його лежить ідея порівняння амплітуд частотних складових аудіосигналу в різних областях спектра. Один з можливих способів визначення спектрального контрасту може виглядати наступним чином:

$SpC = \max(A) - \min(A) / \max(A) + \min(A)$,

де:

- SpC - спектральний контраст;
- A - масив амплітуд частот в аудіосигналі.

Отриманий результат може бути числом в діапазоні від 0 до 1, де 0 означає мінімальний спектральний контраст (всі частоти мають близькі амплітуди), а 1 вказує на максимальний спектральний контраст (є видима велика різниця між амплітудами різних частот). Спектр застосування методу аналізу аудіосигналу за допомогою спектрального контрасту також досить широкий, але основних напрямків дещо менше через вужчу направленість використання.

1. У звуковому дизайні спектральний контраст може бути використаний для створення звукових ефектів або навколишніх шумів, які мають особливий специфічний характер;
2. У галузі розпізнавання мовлення спектральний контраст може служити однією з характеристик, що допомагає в розрізненні різних звуків літер та образів мовлення;
3. В аудіозаписі та синтезі звуку, зокрема й синтезі штучного людського голосу, використання спектрального контрасту дозволяє створювати або аналізувати аудіосигнали з урахуванням контрасту їхньої спектральної складової. Так само, як для синтезу, може бути використаний у зворотному боці для аналізу аудіосигналу.

На рисунку 2.15 можна побачити діаграму спектрального контрасту, яка побудована за допомогою бібліотеки Librosa для мови Python.

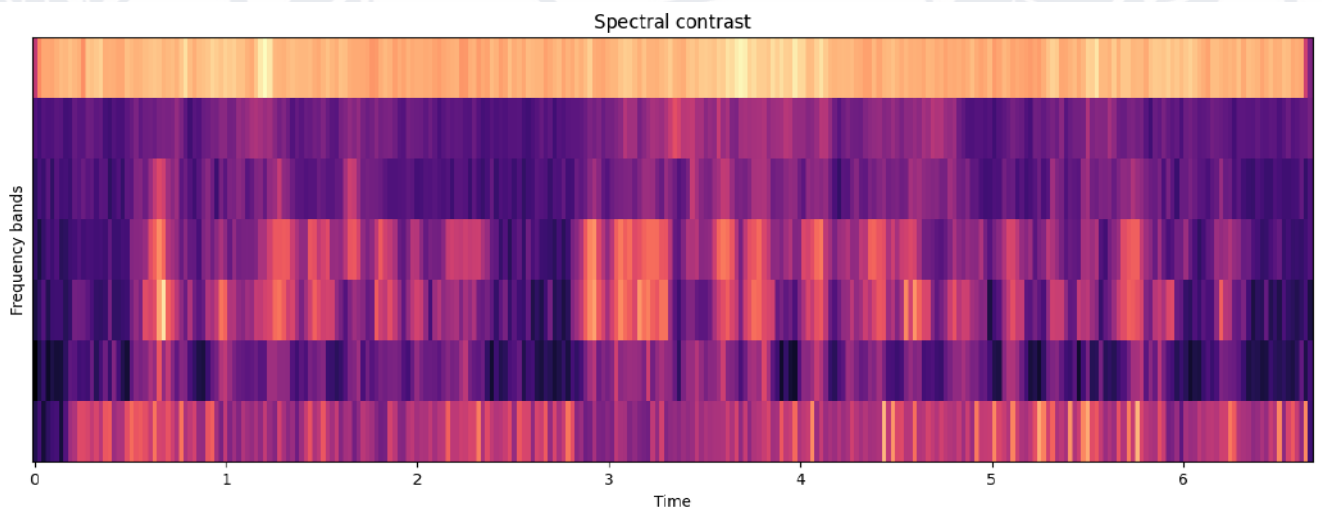


Рисунок 2.15 – Діаграма спектрального контрасту

Хромограма (Chroma Energy Normalized") - Функції кольоровості ґрунтуються на дванадцяти атрибутах написання основного тону, як використовується в західному нотному записі, де кожен вектор кольоровості вказує, як енергія у кадрі сигналу розподіляється по дванадцяти смугах кольоровості (сім основних та п`ять проміжних смуг, що відповідають додатковим

ступеням). Вимірювання таких розподілів у часі складає хромаграму, яка тісно корелює з мелодійною та гармонійною прогресією аудіосигналу. Такі послідовності часто схожі для різних записів. Нормалізована енергія кольоровості застосовується співставлення звуку, де допускаються його варіації, оскільки, зазвичай, породжуються різними людьми, що мають свої унікальні голосові зв`язки. Також одна й та ж людина не може ідентично відтворити двічі один звук аби всі характеристики його аудіосигналу співпали. На рисунку 2.16 можна побачити хромограму, яка побудована за допомогою бібліотеки Librosa для мови Python.

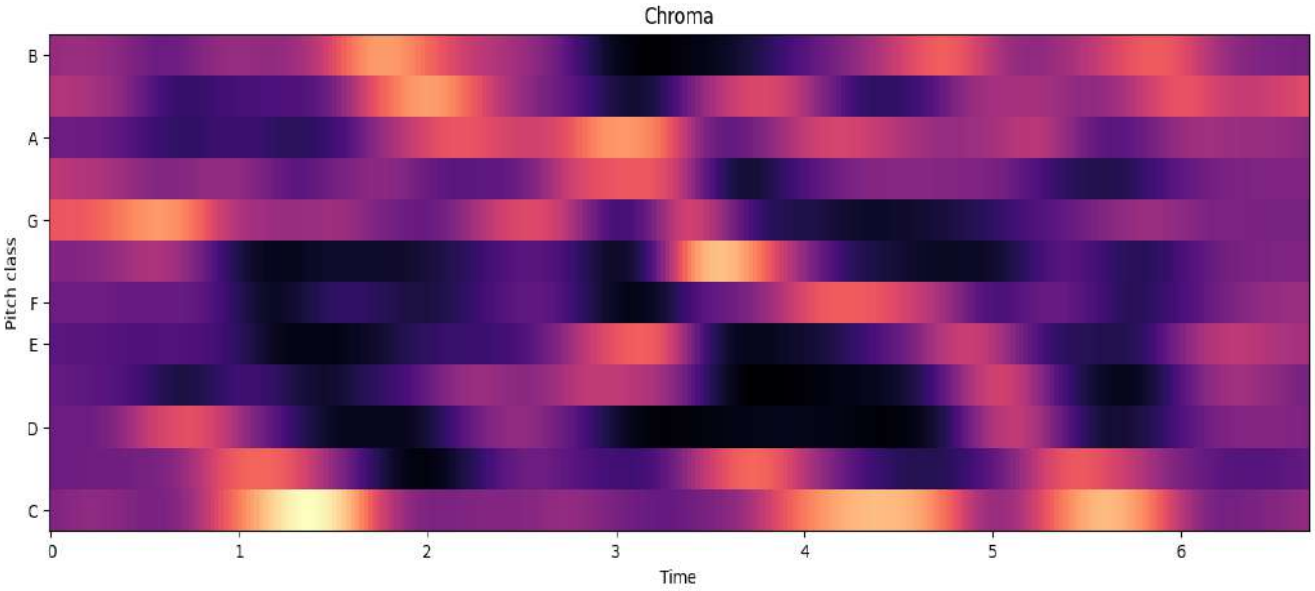


Рисунок 2.16 - Хромограма

Це все – основні характеристики, що потрібно визначити та проаналізувати для проведення широкої перевірки аудіосигналу на предмет присутності на ньому штучно генерованого голосу.

Складність такого методу генерації синтетичного голосу людини полягає у потребі наявності неймовірно великої бази аудіо даних. Адже потрібно мати запис безлічі слів у виконанні живої людини, щоб мати змогу модифікувати їх звучання та зводити їх у великі фрази. Другою складністю є помітність однаковості звучання однакових слів, якщо у базі вони записані лише одним дублем. Це сильно знижує

шанси згенерованого таким чином запису голосу пройти перевірку. Рішенням цієї проблеми може стати запис більшої кількості дублів, що в декілька разів збільшить потребу. Наприклад, середня кількість звуків у слові англійської мови – 5 [26]. А середня тривалість їх вимови становить близько половини секунди. Отже, у хвилині часу поміститься близько 120 слів англійської мови без врахування пунктуації та пауз між ними.

Розмір файлу WAV формату залежить від кількості каналів (моно чи стерео), розширення бітів на зразок, частоти дискретизації та інших параметрів аудіо. Однак я можу надати приблизні розрахунки. Формула для визначення розміру файлу звучання може виглядати приблизно так:

S=Te*B*t*C/8, де:

- S – Розмір;
- Te - Частота дискретизаці;
- B - Розширення бітів;
- t – Тривалість;
- C - Кількість каналів.

Розглянемо приклад для стерео аудіосигналу тривалістю в одну хвилину з розширенням бітів на 24 біт та частотою дискретизації 44,1 кГц, що є стандартом для аудіосигналу у 2023 році:

44100 Гц. * 16 біт * 60 секунд * 2 канали / 8

Розмір=44100Гц×16біт×60секунд/8

Розмір= 10584000 байт

Розмір ≈ 10.6 МБ

Отже, приблизно для стереофонічного аудіо тривалістю 1 хвилину у форматі WAV із стандартними параметрами може займати близько 10.6 мегабайт. Якщо аудіо має вище розширення бітів або частоту дискретизації, розмір файлу може бути більшим. Англійська мова містить у собі понад 273 тис. слів [27]. Отже, за найоптимістичнішими прогнозами, щоб вмістити у себе запис усіх слів потрібно використати:

273000 слів / 120 слів/хв. * 10.6 Мб/хв = 24,115 Гб.

Отже, запис усіх слів однієї особи у чистому вигляді займатиме понад 24 Гб. на жорсткому диску системи. Кількість дублів прямопропорційно збільшить потрібне для цього місце, як і кількість осіб, записи яких проводитимуться. Крім того, безперервне зачитування такої кількості тексту займе близько 38 годин, що ускладнює втілення такого проєкту з технічних причин.

Іншим підходом до урізноманітнення звучання однакових слів є псевдовипадкові зміни, що будуть генеруватися автоматично до початку роботи алгоритму по збору потрібної фрази та обробки її звучання. Цей варіант значно зменшить необхідне місце на жорсткому диску системи, але більше навантажуватиме її апаратну частину через необхідність здійснення додаткових дій.

Все це робить використання змішаного методу генерації синтетичного людського голосу неймовірно складним та доступним лише для неймовірно малої кількості генераторів. Тож для збільшення ефективності роботи системи аналізу потрібно мати можливість проведення поверхневого аналізу, якого буде достатньо для виявлення синтетичного голосу, що згенеровано стандартними методами, а також можливість вибірково провести глибокий аналіз аудіосигналу, де після поверхневої перевірки виникають підозри щодо його походження.

Висновок до розділу 2

У даному розділі було розглянуто існуючі математичні моделі та методи аналізу аудіосигналу. Було оглянуто стандартну універсальну математичну модель аналізу аудіосигналу за допомогою Мел-кепстральних коефіцієнтів, висвітлено принцип її роботи. Було детально розглянуто основні інструменти аналізу аудіосигналу, та висвітлена специфіка реагування людського вуха за графіком гучності Флетчера-Мунсона а також інструментів аналізу на інструментів аналізу на аудіосигну. Визначено ключові характеристики для генерації штучного голосу, що дає розуміння основних актуальних аспектів аналізу аудіосигналу. Розглянуто специфіку змішаного методу генерації синтетичного голосу, виділені основні тези та кроки його генерації, проведено їх аналіз, та запропоновано методику

альтернативну методику глибокого аналізу аудіосигналу, що включає широкий аналіз його спектральних характеристик. Проведено огляд інструментів глибокого спектрального аналізу аудіо даних. Розглянуто галузі їх використання, а також специфіку та принципи їх роботи. Наведено приклади візуалізації їх роботи. Висвітлено технічні складнощі та нюанси, що можуть виникнути в процесі генерації змішаного типу. На основі цього обґрунтовано необхідність створення можливості проведення глибокого спектрального аналізу аудіосигналу лише для конкретних випадків з метою заощадження апаратних потужностей аналізуючої системи, а також збільшення її швидкодії. У наступному розділі буде проведено огляд програмного забезпечення, що було обране для виконання поставленої задачі.

РОЗДІЛ 3
ОГЛЯД ТЕХНОЛОГІЙ ДЛЯ СТВОРЕННЯ ПРОГРАМНОГО ПРОДУКТУ

3.1 PyCharm

PyCharm – це інтегроване середовище розробки (IDE), що засноване на програмному забезпеченні IntelliJ IDEA від компанії JetBrains – офіційний засіб розробки додатків на мові Python (рисунок 3.1).

PyCharm був випущений на ринок інтегрованих середовищ розробки для Python щоб створити конкуренцію з PyDev і поширенішим середовищем розробки Komodo IDE. Бета-версія була випущена в липні 2010, версія 1.0 була випущена трьома місяцями пізніше.

PyCharm Community Edition, безкоштовна версія з відкритим початковим кодом була опублікована 22 жовтня 2013.

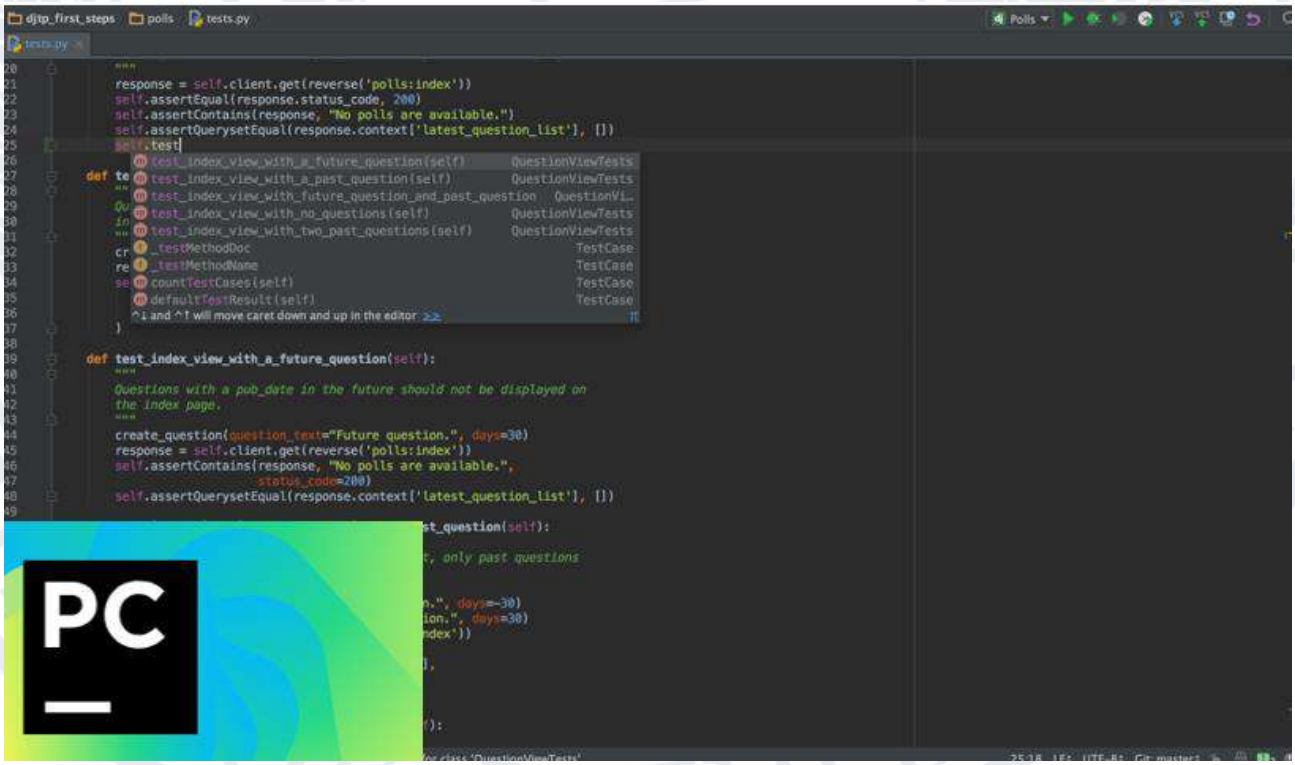


Рисунок 3.1 – Інтегроване середовище розробки PyCharm

Це інтегроване середовище розробки надає велику кількість можливостей, основні з яких:

1. Статичний аналіз коду, підсвічування синтаксису і помилок;
2. Навігація серед проєктів і початкового коду: відображення файлової структури проєкту, швидкий перехід між файлами, класами і методами;
3. Рефакторинг: перейменування, витяг методу, введення змінної, введення константи, підняття і опускання методу тощо;
4. Інструменти для веброзробки з використанням фреймворку Django;
5. Вбудований зневаджувач для Python;
6. Вбудовані інструменти для юніт-тестування;
7. Розробка з використанням Google App Engine;
8. Підтримка систем контролю версій: загальний користувацький інтерфейс для Mercurial, Git, Subversion, Perforce і CVS з підтримкою списків змін та злиття.

Переваги PyCharm:

1. Інтегроване середовище розробки (IDE): PyCharm надає повноцінне інтегроване середовище розробки, спеціально адаптоване для мови програмування Python, що включає в себе вбудовану консоль, дебагер, редактор коду та інші корисні інструменти;
2. Підтримка багатьох фреймворків: PyCharm підтримує багато популярних фреймворків для розробки на Python, таких як Django, Flask, PyTest, NumPy, і багато інших;
3. Розумний автодоповнювач: Має потужний і розумний автодоповнювач, який полегшує написання коду та допомагає у виявленні помилок ще на етапі написання коду;
4. Система контролю версій: Інтегрована підтримка систем контролю версій, таких як Git, дозволяє легко взаємодіяти з кодом у командному середовищі;

5. Аналіз коду та виявлення помилок: Вбудовані інструменти для аналізу коду дозволяють виявляти потенційні проблеми та помилки ще до виконання програми;
6. Рефакторинг: PyCharm має ряд інструментів для автоматизованого рефакторингу, що полегшує покращення структури та читабельності коду.
7. Підтримка віртуальних середовищ та пакетний менеджмент: Інтегрована підтримка віртуальних середовищ (наприклад, virtualenv) та пакетний менеджмент (pip) дозволяє легко керувати залежностями потрібного проєкту.
8. Крос-платформеність: PyCharm підтримує операційні системи Windows, macOS та Linux, що робить його доступним для багатьох розробників, які надають перевагу тим, чи іншим операційним системам.

Недоліки PyCharm

PyCharm є фактично ідеальним IDE для створення додатків написаних на мові Python, але все ж має незначну кількість недоліків через свою специфіку та напрямленість:

5. Для комфортної роботи в середовищі знадобиться потужний комп’ютер. Через наявність великої кількості додаткових зручностей у цьому IDE, воно автоматично сильно навантажує систему, що може стати проблемою для комп`ютерів з комплектуючими від десятирічної давнини і більше;
6. Іноді можливі проблеми при першому розгортанні проєктів, отриманих від інших розробників;
7. Час завантаження: Час завантаження IDE може бути трошки довгим, зокрема, при великих проєктах;
8. Обмеження у спільнотській версії: Деякі функції та інструменти можуть бути обмежені у безкоштовній (Community) версії PyCharm, що може обмежити можливості деяких розробників.

Попри ці недоліки, PyCharm залишається однією з найпопулярніших та ефективних IDE для розробки на мові програмування Python, завдяки своїм розширеним можливостям та інструментам, які сприяють швидкому та продуктивному процесу розробки.

3.2 Мова програмування Python

Python — інтерпретована об'єктно-орієнтована мова програмування високого рівня зі строгою динамічною типізацією. Розроблена в 1990 році Гвідо ван Россумом. Структури даних високого рівня разом із динамічною семантикою та динамічним зв'язуванням роблять її привабливою для швидкої розробки програм, а також як засіб поєднування наявних компонентів.

Python підтримує модулі та пакети модулів, що сприяє модульності та повторному використанню коду. Інтерпретатор Python та стандартні бібліотеки доступні як у скомпільованій, так і у вихідній формі на всіх основних платформах.

В мові програмування Python підтримується кілька парадигм програмування, зокрема: об'єктно-орієнтована, процедурна, функціональна та аспектно-орієнтована [28].

Як і будь-які інші мови програмування, мова Python має свої переваги та недоліки.

Переваги мови Python:

1. Досить сучасний мовний синтаксис і можливість використання різноманітних парадигм програмування (об’єктно-орієнтована, функціональна);
2. Дружнє ком’юніті, велика кількість відповідей на запитання на StackOverflow, що робить поріг входження набагато нижчим;
3. Python йде в пакеті «з батареями», як говорять пайтоністи, тобто з дуже великою бібліотекою стандартних пакетів (не треба їх додатково інсталювати, щоб створювати навіть досить складні проєкти) та ще більшим списком пакетів Open Source, доступних в публічних репозиторіях (для легкого встановлення). У тому числі потужні пакети для

обробки зображення, створення графіків, статистичного аналізу даних, мережевої комунікації, штучного інтелекту (особливо Deep Learning) та Machine Learning, опрацювання натуральної мови, взаємодії з базами даних, створення вебдодатків, витягування даних з вебсайтів, графічних алгоритмів;

4. Незалежність від платформи — програми в Python працюють здебільшого на кожній платформі, для якої доступний інтерпретатор Python. Вистачає раз написати й можна запускати на майже кожній операційній системі;
5. Python є чудовою мовою програмування для прототипування, тобто швидкого написання пробних версій додатка, які мають перевірити підхід або довести, що якась концепція спрацює. А це завдяки стислому й багатому синтаксису, вищепереліченій насиченості пакетів і мінімальним вимогам для початку створення додатка;
6. Величезна кількість матеріалів у мережі, чудова документація для великої кількості стандартних пакетів;
7. Керована пам’ять — інтерпретатор Python сам звільняє пам’ять, виділену програмі, але яка вже не використовується. Це полегшує продумування програми, скорочує код і унеможливлює допускання помилок, пов’язаних із передчасним звільненням ще потрібної пам’яті, які часто з’являються в програмах, написаних, наприклад, мовами C або C++.

Недоліки мови Python:

1. Складна система публікування власних пакетів Open Source, особливо, якщо використовують бібліотеки в C — Python, на жаль, стає тут жертвою власного успіху й успадковує труднощі, пов’язані з побудовою нативного коду (наприклад, мовою C) на конкретну системно-апаратну платформу;
2. Система типізації — це одночасно перевага й недолік. Для програмістів-початківців певним шоком може бути те, що в програмі не треба подавати типи для змінних. Python розпізнає їх сам й відповідно перевіряє в ході діяльності програми, чи ми не пробуємо виконати на даних недозволені

операції (так звана сильна система типізації). Однак це відбувається лише на етапі виконання конкретного шматка коду, а тому про потенційну помилку дізнаємося лише після запуску. Це зумовлює те, що, створюючи програми на Python, більшу увагу слід приділити їхньому тестуванню. На відміну від компільованих мов програмування (C, C++, Java, C# та багато інших), компілятор не допоможе нам викрити певні помилки, тому ми повинні самі подбати про відповідне покриття коду тестуваннями. Відсутність явної типізації також є певного роду ускладненням, коли ми розширюємо велику систему. Тоді явна типізація є важливою допомогою для програміста, який читає код, написаний кимось іншим. Тому також нові версії мови Python мають опційну можливість опису функцій і класів через типи, що є рекомендованою практикою у випадку більших додатків;

3. Зважаючи на широку екосистему й багатство залежностей, певні виклики дає контейнеризація додатка в Python, а точніше — управління залежностями й будування образів додатка, які мали б використовуватися в сконтейнеризованому середовищі (наприклад, докер). Це найбільш можливо, але треба уважно контролювати особливо пакети, які використовують нативні бібліотеки.
4. Окремим недоліком можна виділити невелику швидкодію мови Python. Однак, ця проблема вирішується за допомогою використання популярної Python бібліотеки NumPy, що використовує код більш швидкої мови C++ для реалізації методів Python [29].

3.3Librosa

Бібліотека Librosa була розроблена та написана для мови програмування Python і вперше випущена в 2011 році, але залишається актуальною та отримує патчі кожні декілька місяців (рисунок 3.2). Librosa - це потужна бібліотека мови програмування Python, спеціально розроблена для аналізу аудіосигналів та взаємодії з ними. Бібліотека дуже проста у використанні та може виконувати як базові, так і складні завдання, пов'язані з обробкою аудіо та музики [30].

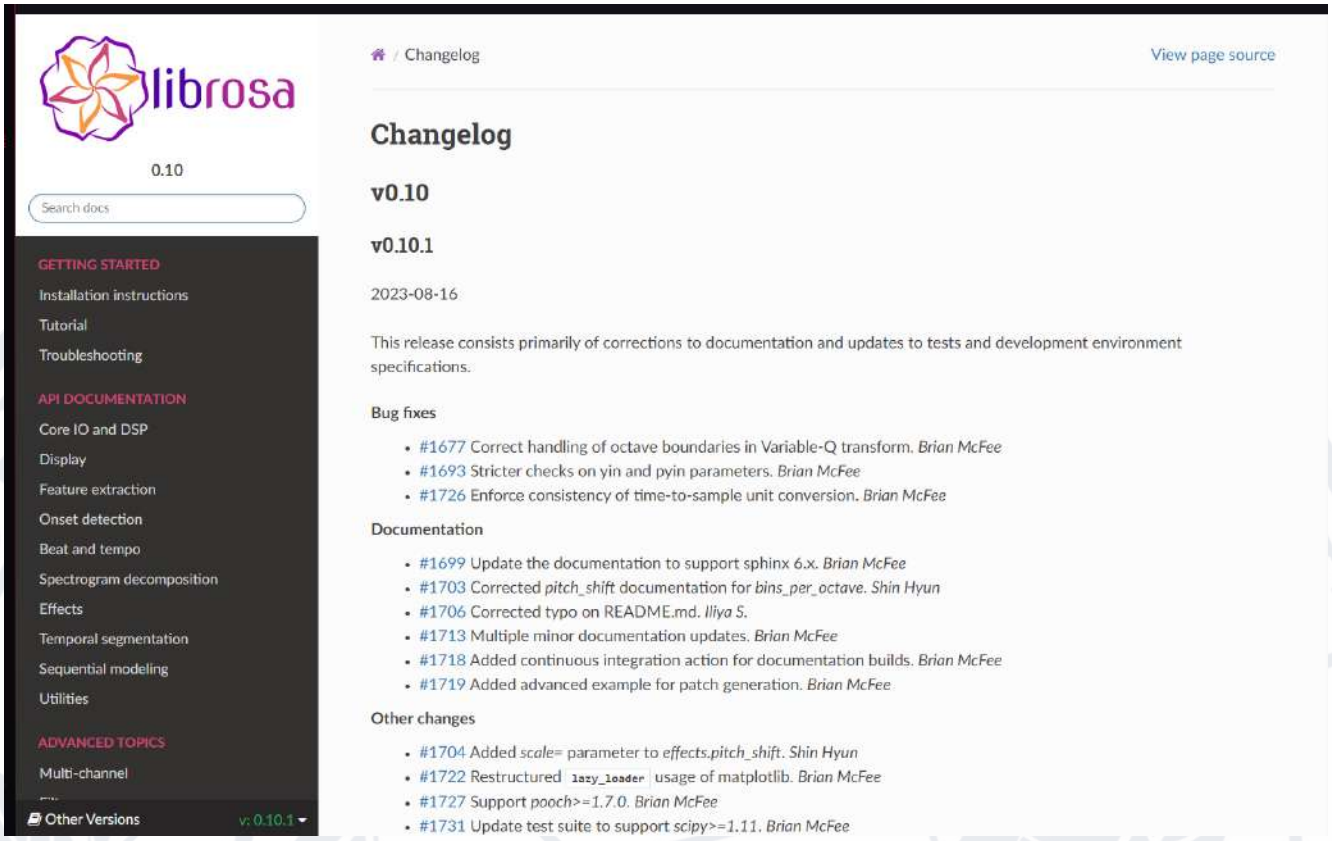


Рисунок 3.2 – Сторінка Changelog на офіційному сайті Librosa

Бібліотека має відкритий вихідний код та знаходиться у вільному доступі за ліцензією ISC. Основним завданням Librosa є надання інструментів для екстракції корисної інформації з аудіоданих, що використовується у багатьох областях, таких як обробка мови, музичний аналіз, машинне навчання і багато інших. Ось кілька основних можливостей бібліотеки Librosa:

1. Завантаження аудіо у проєкт: Librosa дозволяє легко завантажувати аудіофайли в різних форматах, таких як WAV, MP3, MP4, M4a, Flac, OGG, тощо. Вона надає функції для читання аудіофайлів з носія та отримання аудіоданих та розмірностей;
2. Екстракція характеристик аудіосигналу: Librosa дозволяє витягувати різні характеристики з аудіосигналів, такі як мел-фреквенційні кепстральні коефіцієнти (MFCC), спектрограми, хромаграми, темп та інші. Ці характеристики можуть використовуватись для подальшого аналізу або для подачі на вхід у моделі машинного навчання;

3. Робота з ритмом та темпом: Librosa має функції для виявлення ритму та вимірювання темпу аудіосигналу;
4. Спектральний аналіз: Librosa надає функції для обчислення спектрограм, стфт, частотних характеристик та інших спектральних властивостей аудіосигналу.

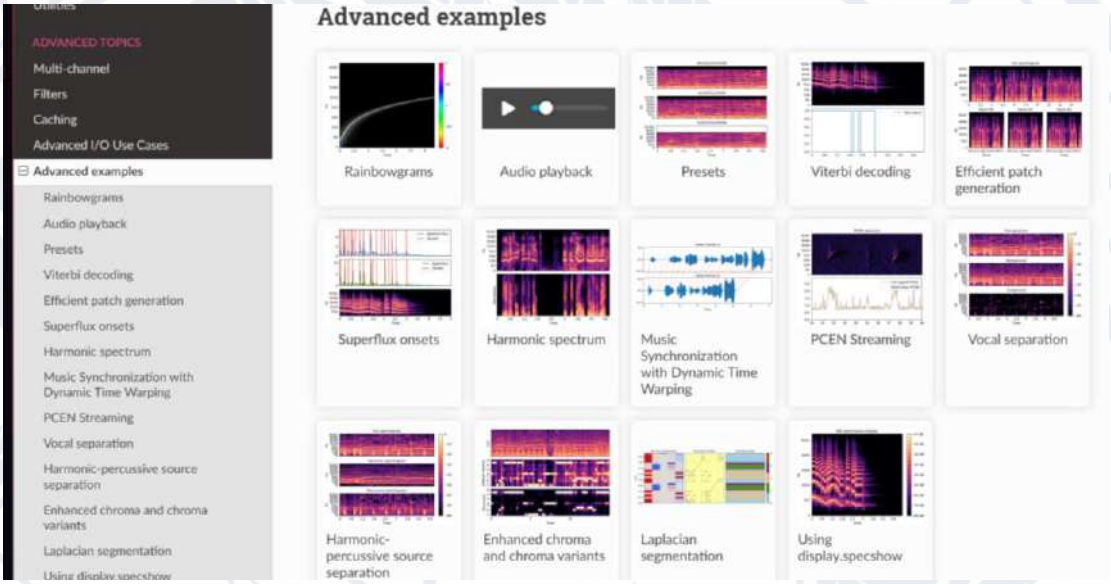


Рисунок 3.3 – Розділ офіційного сайту з прикладами використання функцій бібліотеки

Як і будь-яке інше програмне забезпечення, бібліотека аналізу аудіо даних Librosa має свої переваги та недоліки. Для того, щоб визначити чи актуальним буде її використання для виконання поставленої задачі, потрібно провести їх аналіз, та підбити підсумки.

Переваги бібліотеки Librosa:

1. Висока функціональність: Librosa надає широкий спектр інструментів для аналізу аудіосигналів, включаючи створення спектрограм, хромограм, аналіз тембру, визначення ритму та інші корисні функції;
2. Простота використання: Легкий інтерфейс бібліотеки дозволяє вам швидко та ефективно працювати з аудіоданими, що особливо важливо для початківців;

3. Активна спільнота та документація: Є докладна документація та активна спільнота користувачів, що полегшує вирішення проблем, обмін досвідом та отримання підтримки;
4. Окрім наявності детальної великої документації для ознайомлення з принципами роботи бібліотека, можна виділити наявність на сайті бази даних з прикладами використання різних функцій бібліотеки для зниження вхідного порогу для роботи з бібліотекою (рисунок 3.3).
5. Широке застосування: Librosa знаходить застосування в різних галузях, таких як обробка мови, музичний аналіз, машинне навчання і дослідження звукових сигналів;
6. Регулярні оновлення: Бібліотека регулярно оновлюється, додаючи нові функції та виправляючи помилки, що сприяє її розвитку та залишає її у числі актуальних для використання бібліотек;
7. Librosa також має інші функції для аналізу аудіосигналів, такі як розпізнавання тембру, знаходження пік-фактору та інших артефактів, присутніх на записі, тощо. Вона є потужним інструментом для витягування характеристик та аналізу аудіоданих у Python.

Недоліки бібліотеки Librosa:

1. Швидкодія: У випадках обробки великих обсягів аудіоданих, Librosa може виявитися менш ефективною з точки зору швидкодії порівняно з деякими іншими бібліотеками. Дію бібліотеки можна пришвидшити за допомогою використання її можливостей співпраці з іншими бібліотеками-оптимізаторами, код яких в більшості написаний на швидкій мові C++. Але це породжує наступний недолік роботи з бібліотекою аналізу аудіоданих Librosa;
2. Залежність від інших бібліотек: Деякі функції для пришвидшення процесу виконання завдання можуть вимагати наявності додаткових бібліотек, таких як NumPy, SciPy та інших, що може становити проблему у деяких середовищах;

3. Неінтуїтивність для новачків: Для тих, хто тільки починає вивчати обробку сигналів, деякі аспекти інтерфейсу Librosa можуть здаватися неінтуїтивними або вимагати додаткового часу для освоєння;
4. Обмеженість у роботі з великими обсягами даних: В деяких випадках, при обробці великих аудіофайлів, може виникати обмеженість з точки зору обсягу доступної оперативної пам'яті;
5. Оновлення та зміни в API: Іноді оновлення та зміни в API можуть призводити до несумісності з попередніми версіями коду, що вимагає внесення змін у вже наявний код.

Не зважаючи на ці недоліки, Librosa залишається важливим інструментом для аналізу аудіосигналів у Python, забезпечуючи розширені можливості для дослідження та розробки у галузі обробки сигналів та аналізу музики.

Висновок до розділу 3

У даному розділі перелічені інструменти, використані для виконання поставленої задачі. У наступному розділі буде розглядатися власне програмне рішення поставленої задачі.

РОЗДІЛ 4
ОГЛЯД ВЛАСНОГО МЕТОДУ АНАЛІЗУ АУДІОСИГНАЛУ ТА СТВОРЕННЯ ПРОГРАМНОГО ПРОДУКТУ

Для демонстрації роботи власного алгоритму аналізу аудіосигналу було розроблено програмне забезпечення. Це не є повноцінним програмним забезпеченням для ідентифікації синтетичного голосу, а тільки фрагмент визначення ключових характеристик, обумовлений у рамках даної роботи.

4.1 Опис розробленої програми

Розроблене ПЗ є демонстраційним додатком для виокремлення та обрахунку ключових характеристик аудіосигналу, а також їх візуалізації для подальшого їх використання. Під час розробки програмного забезпечення крім бібліотеки librosa були використані й інші, такі як: IPython, numpy, pandas, scipy, matplotlib, а також seaborn. Більшість з них потрібні для коректної роботи бібліотеки аналізу аудіоданих. У програмі було інтегровано власний покращений блок ідентифікації параметрів. Він працює наступним чином: після того, як програма отримала шлях до потрібного для аналізу файлу, що містить певні аудіодані, вона завантажує файл та викликає із бібліотеки librosa функцію перетворення аудіосигналу на звукові хвилі, після чого записує дані у масив, на основі якого будує графік звукових хвиль.

Після того, як дані стосовно звукових хвиль аудіосигналу отримані, програма викликає функцію, що проводить розклад звукових хвиль на гармонічні та ударні хвилі. Ідентично попередньому кроку, програма формує масив елементів, та на його основі будує графік розділу гармонічних та перкусійних хвиль голосу. На рисунках 4.1, 4.2, 4.3 та 4.4 наведено приклад побудови схеми звукових хвиль, та розкладання їх на гармонічні та ударні сигнали, що розділяються на дві різні хвилі. Приклади наведено для тесту природного та штучного голосу.

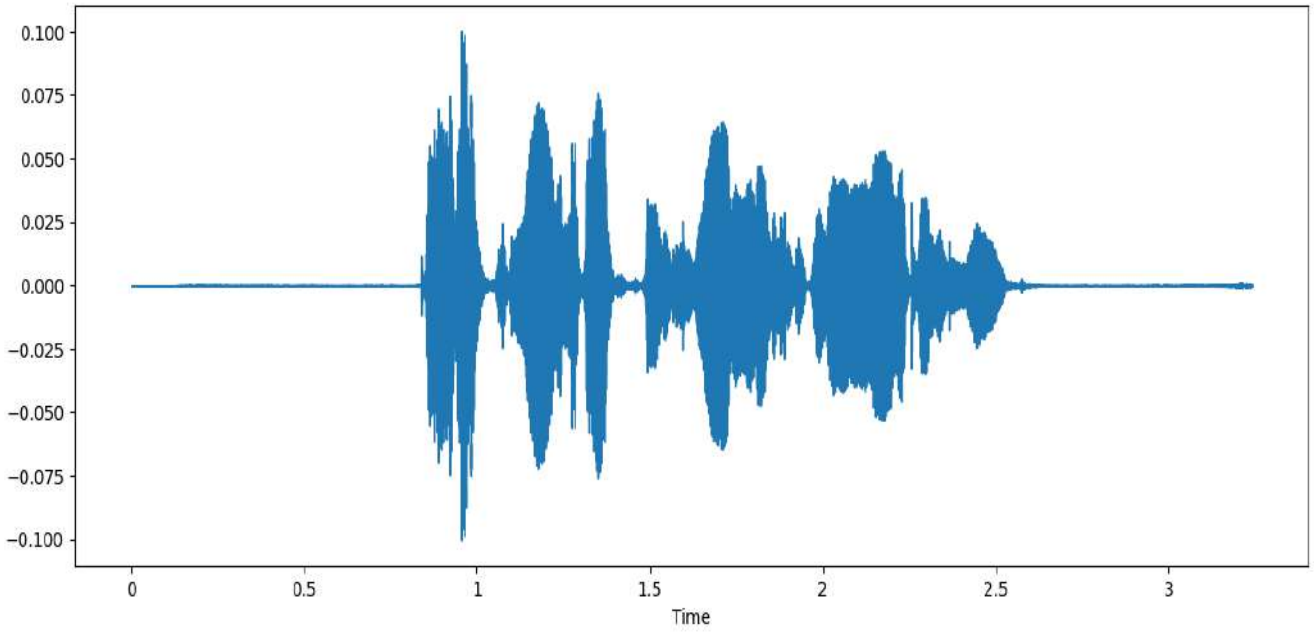


Рисунок 4.1 – Звукові хвилі природного голосу

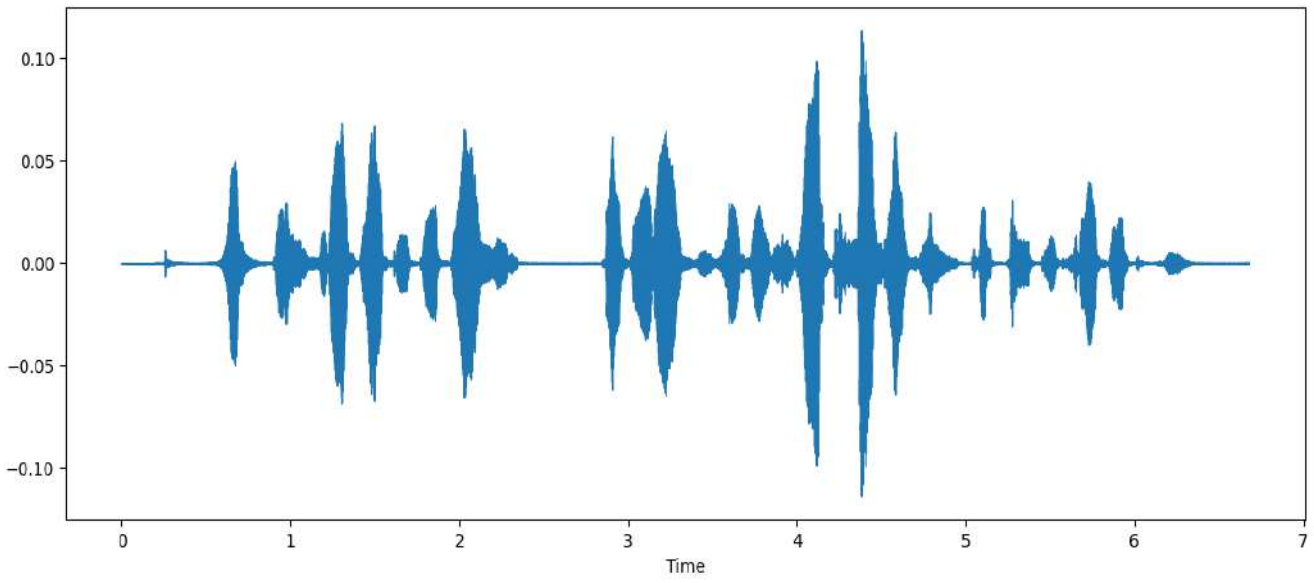


Рисунок 4.2 – Звукові хвилі штучного голосу

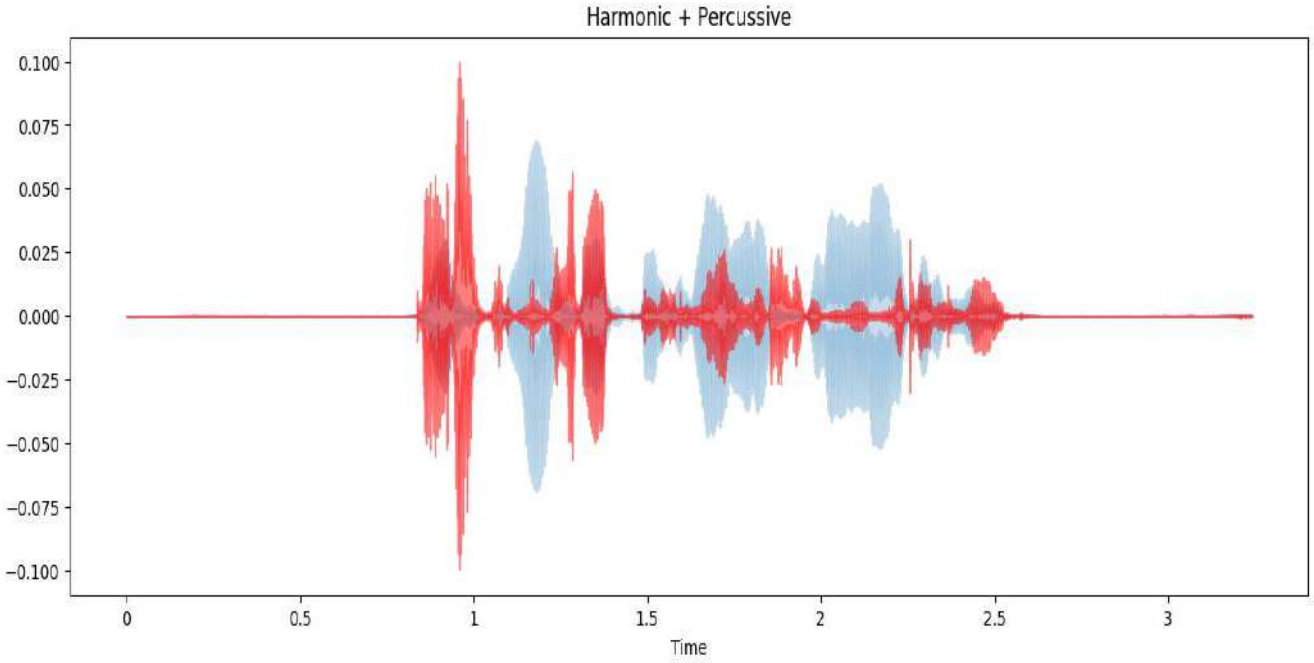


Рисунок 4.3 – Розділ гармонічних та ударних хвиль природного голосу

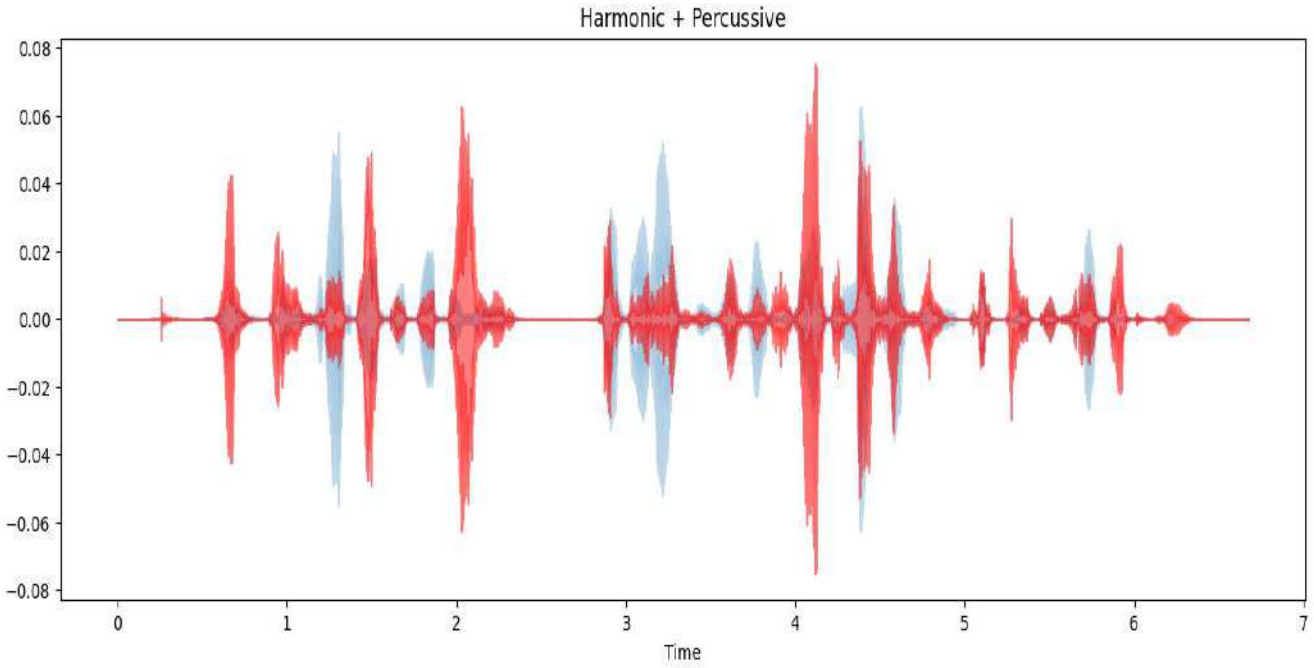


Рисунок 4.4 – Розділ гармонічних та ударних хвиль штучного голосу

Після успішного перетворення ауліосигналу на графік звукових хвиль та графік розподілу гармонічних та перкусійних хвиль, програма, оперуючи отримані дані, проводить базовий спектральний аналіз та утворює спектрограму цього аудіосигналу (рис 4.5, 4.6) аби отримати чіткі дані про гучність звуку на

різних відрізках часу та плавність його зміни. Як можна помітити на рисунках, спектрограма природного голосу має чіткий та яскравий відтінок навіть на висоті, що не властиві людському голосу, а саме на відрізку частот від 20 до 60 Гц, а також від 2кГц і вище. Все це зумовлено людською анатомією та резонансом природніх частот всередині гортані та ротової порожнини людини. Спектрограма синтетичного голосу навпаки, має чітко визначений діапазон, та майже не має ознак сигналу на рівні від 1.5-2кГц і вище, що свідчить про синтетичну природу голосу.

Також, одразу після проведення базового спектрального аналізу, програма викликає функцію обчислення значень Мел-кепстрального коефіцієнту, після чого на основі цих даних будує його графік (рис 4.7, 4.8). Оскільки значення Мел-кепстральних коефіцієнтів є значенням умовної, складеної величини, що відповідає за симуляцію сприйняття сигналу людським вухом, їх візуалізація не дає чіткого розуміння природи аудіосигналу, присутнього на записі. Тому ці множина цих значень призначена виключно для подальшої її обробки нейромережею, що співставить їх з власними еталонами.

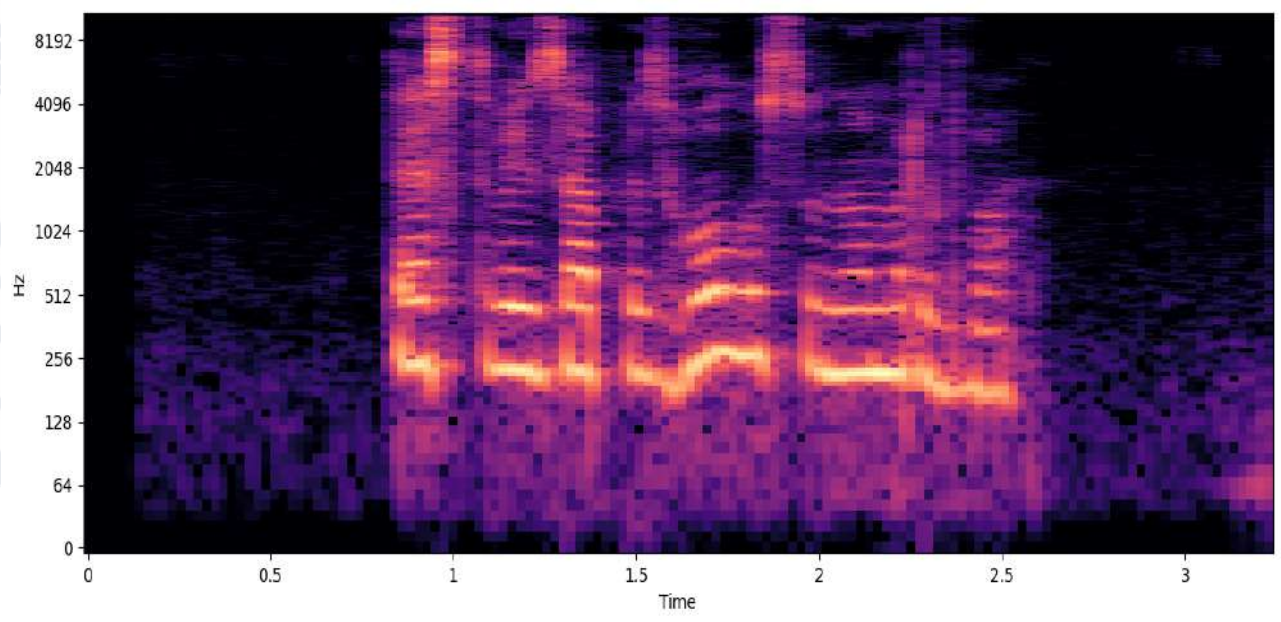


Рисунок 4.5 – Спектрограма природного голосу

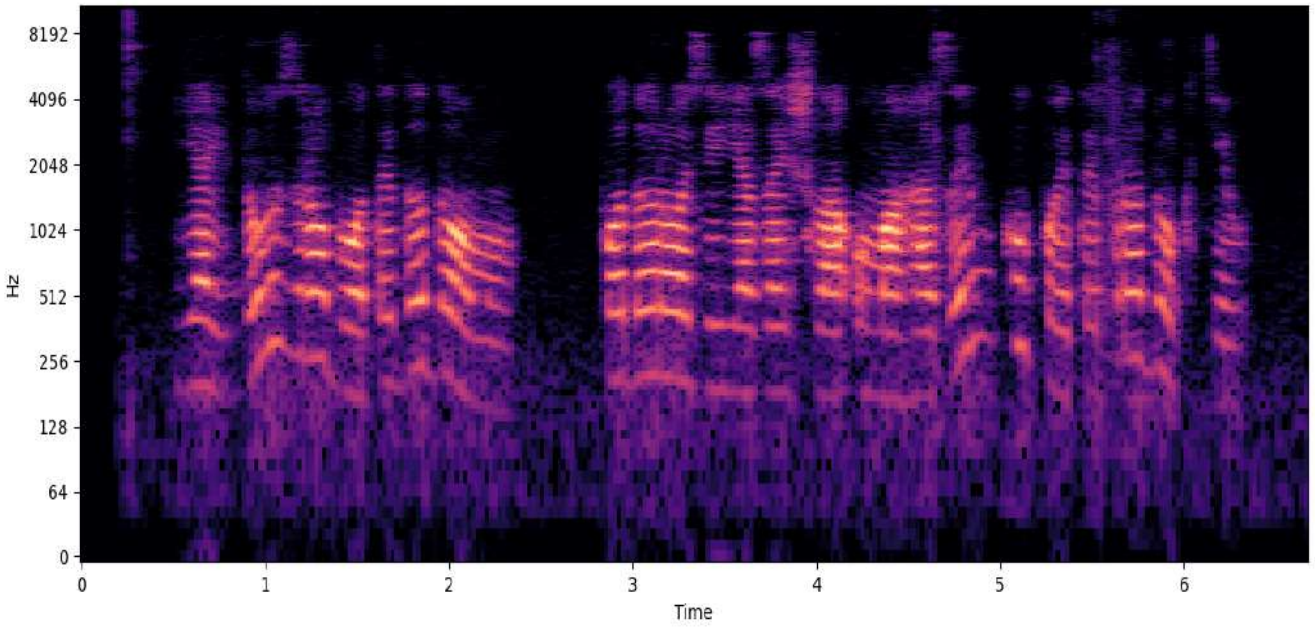


Рисунок 4.6 – Спектрограма штучного голосу

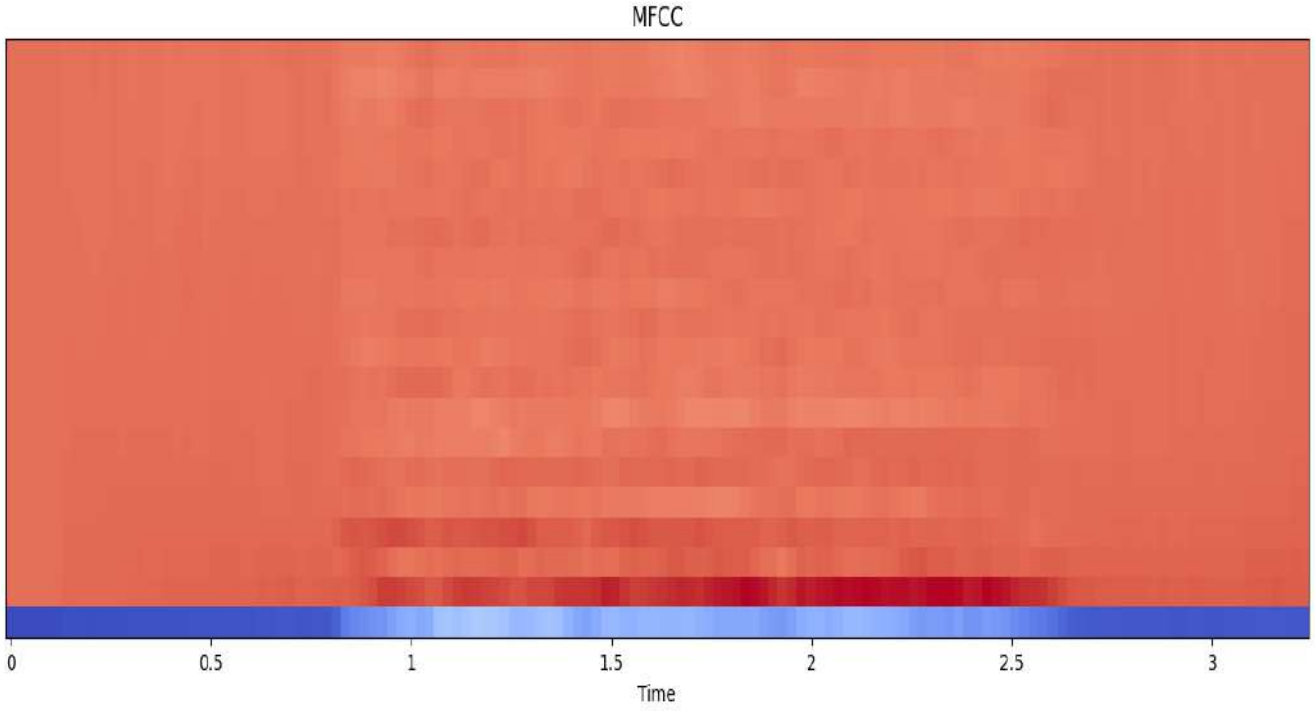


Рисунок 4.7 – Діаграма Мел-кепстральних коефіцієнтів природного голосу

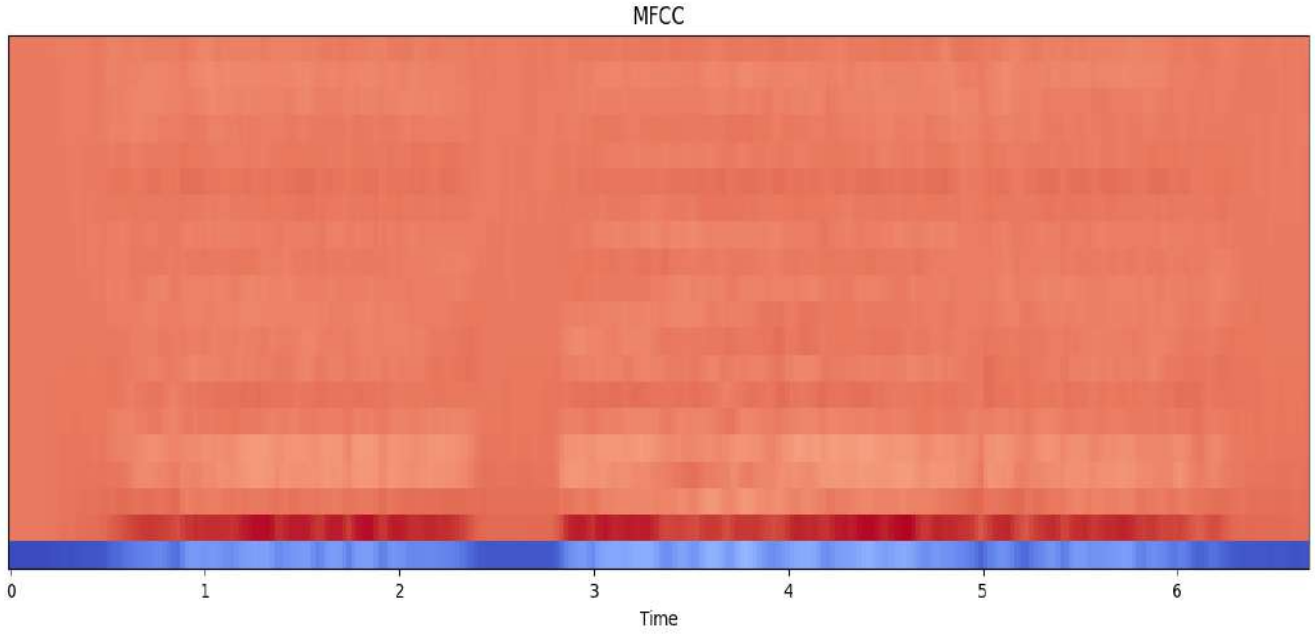


Рисунок 4.8 – Діаграма Мел-кепстральних коефіцієнтів штучного голосу

При аналізі аудіосигналу, на якому потенційно присутній синтетичний голос, згенерований змішаним методом, є актуальним проведення обчислень інших коефіцієнтів, які дозволять ідентифікувати його, зважаючи на специфіку процесу його генерації. Для проведення більш глибокого аналізу аудіосигналу програма діє наступним чином:

1. На основі даних, що були отримані під час проведення базового аналізу аудіосигналу, програма визначає його хромограму, що дозволяє виявити певні патерни тональних перепадів на записі, що характерні для синтетичного голосу, згенерованого змішаним методом;
2. Після проведення хроматичного аналізу аудіосигналу програма проводить аналіз спектрального контрасту, що в купі з результатами хроматичного аналізу дозволяє краще розкрити динаміку спектральних змін звуку, та точніше ідентифікувати наявність або відсутність у ньому синтетичного голосу;
3. Наступним кроком є обчислення спектрального центроїду аудіосигналу, та знаходження його локальних центрів мас. На основі цих даних наступні блоки мають оцінити їх значення та співставити їх з еталонами

для отримання розуміння динаміки потужності аудіосигналу на певних спектральних відрізках та ідентифікації синтетичного голосу у разі знаходження патернів, що не є характерними для природнього голосу людини;

4. Кінцевим кроком покращеного блоку ідентифікації параметрів є обчислення значень спектрального спаду аудіосигналу, що дає можливість точніше ідентифікувати наявність синтетичного голосу на сигналі у разі застосування на етапі його генерації спектральних виправлень методом зведення до певного середнього значення кінця однієї частини та початку другої частини запису.

Отже, для проведення глибокого спектрального аналізу аудіосигналу використано декілька інструментів. На рисунках 4.9, 4.10, 4.11, 4.12, 4.13, 4.14, 4.15, 4.16 наведено хромограми, графіки спектрального контрасту, спектрального центроїду та центрального спаду для природного голосу, та синтетичного голосу, створеного змішаним методом. На рисунку 4.17 наведена схема алгоритму покращеного блоку визначення ключових параметрів аудіосигналу.

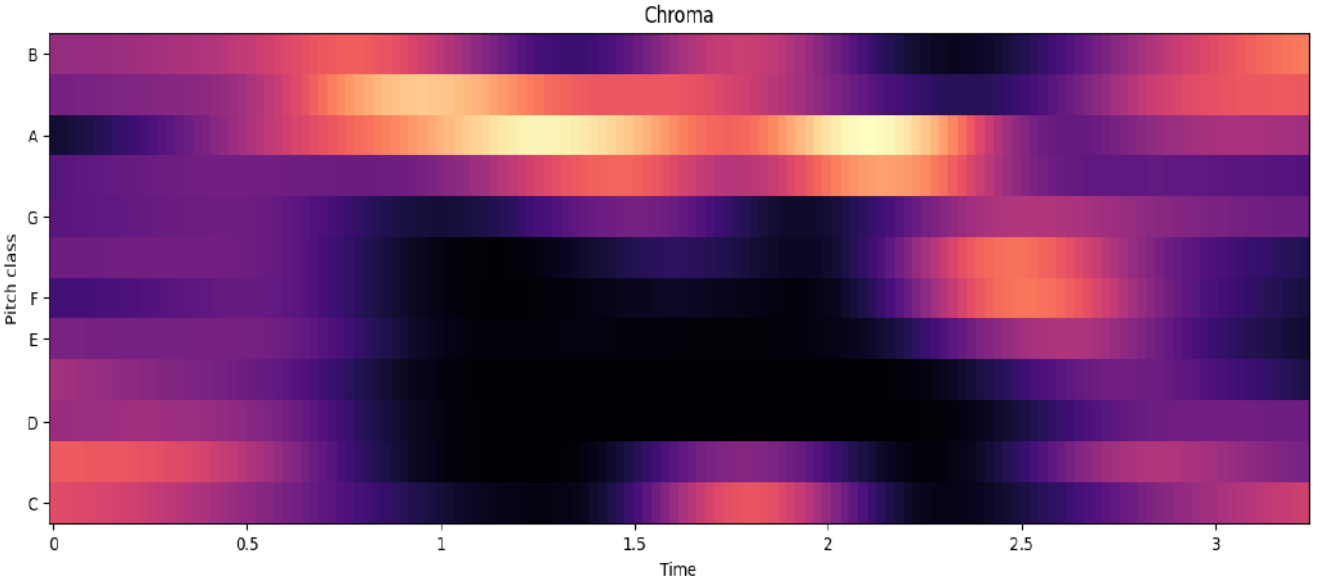


Рисунок 4.9 – Хромограма природного голосу

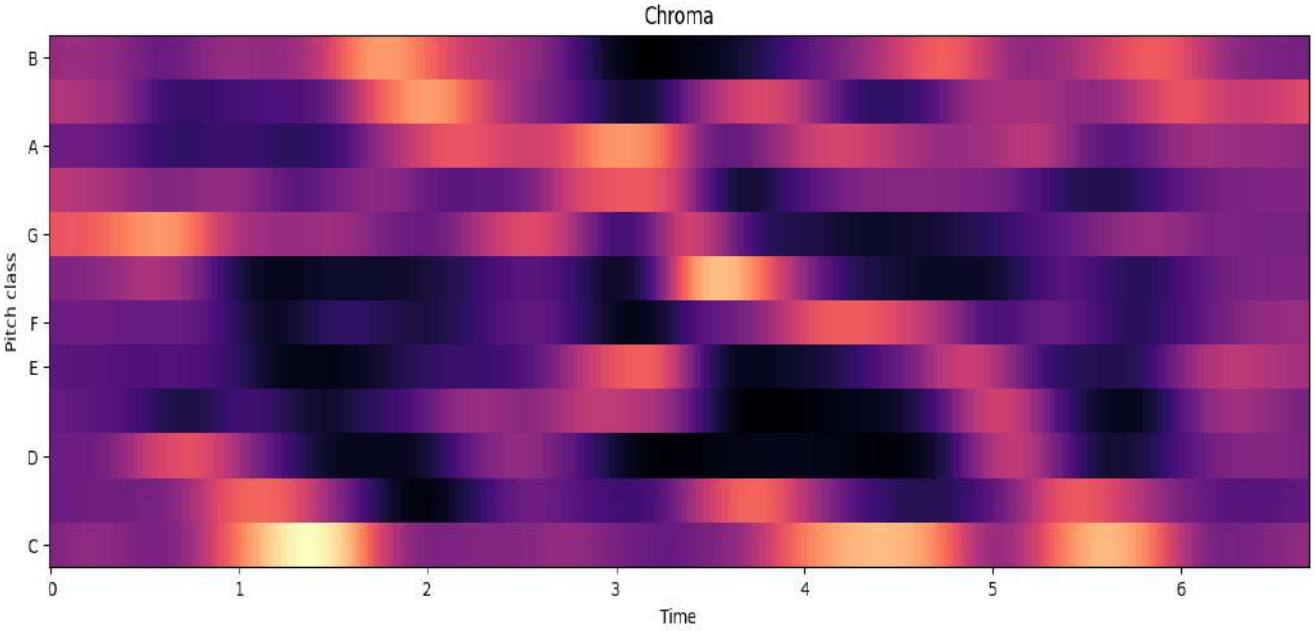


Рисунок 4.10 – Хромограма штучного голосу

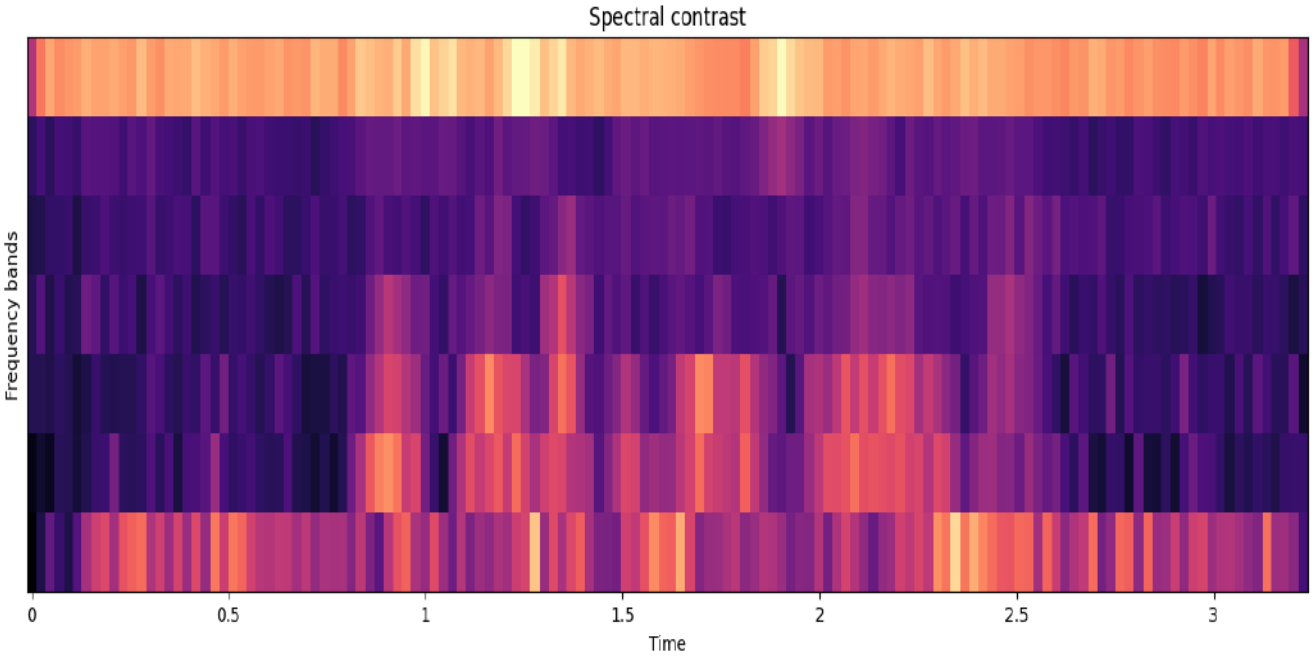


Рисунок 4.11 – Діаграма спектрального контрасту природного голосу

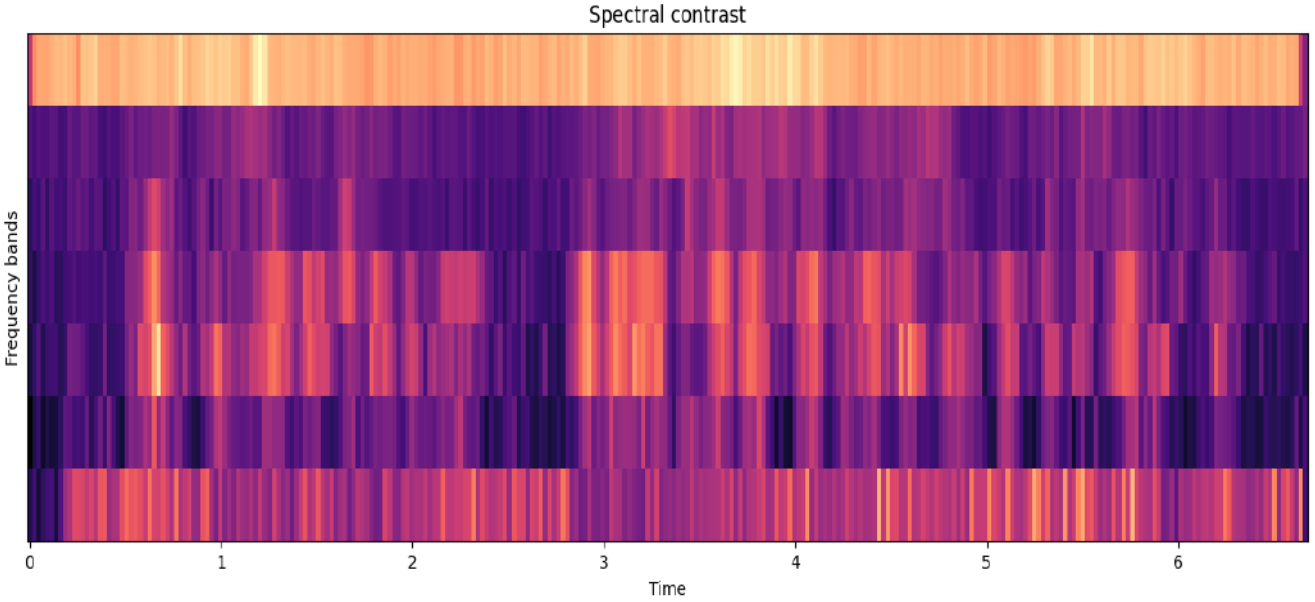


Рисунок 4.12 – Діаграма спектрального контрасту штучного голосу

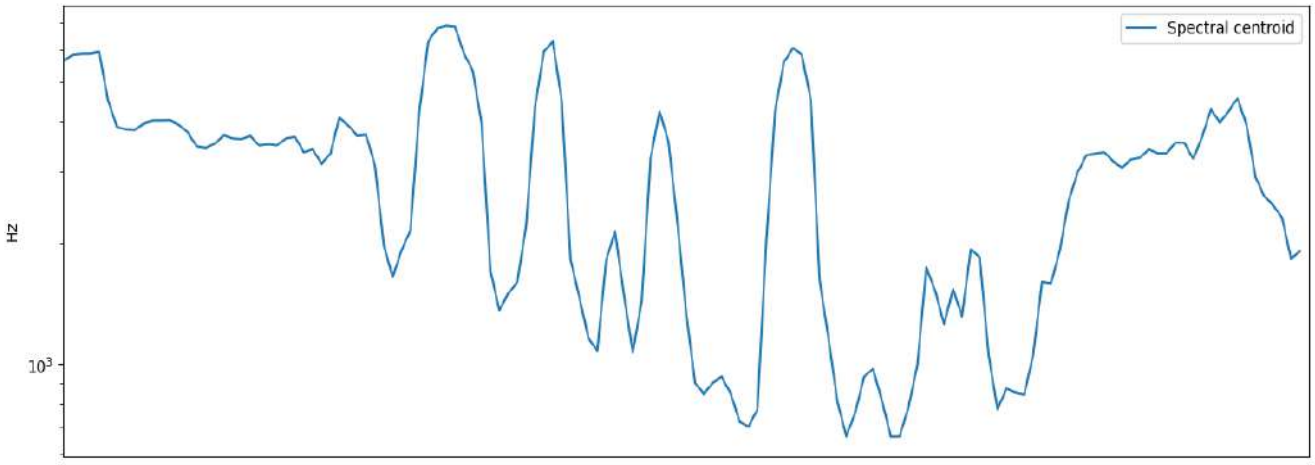


Рисунок 4.13 – Діаграма спектрального центроїду природного голосу

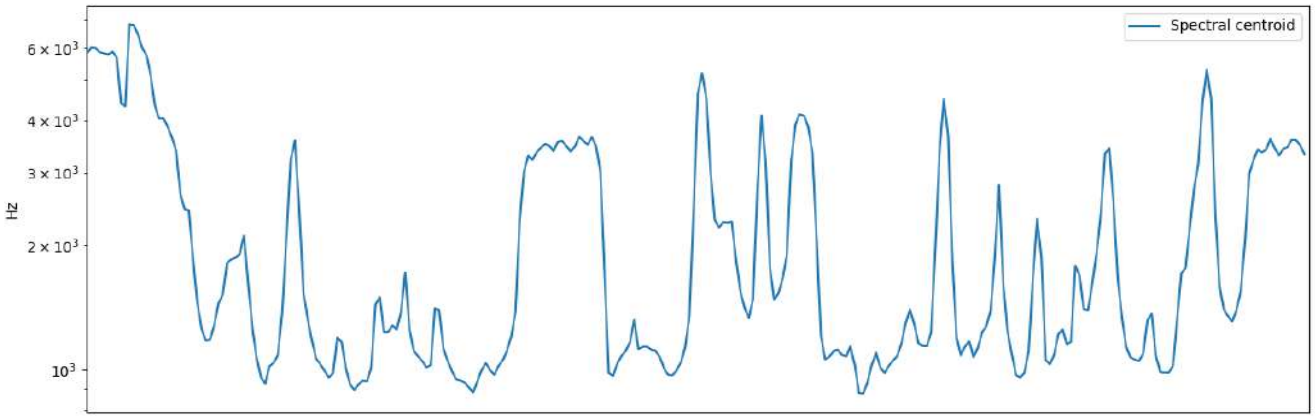


Рисунок 4.14 – Діаграма спектрального центроїду штучного голосу

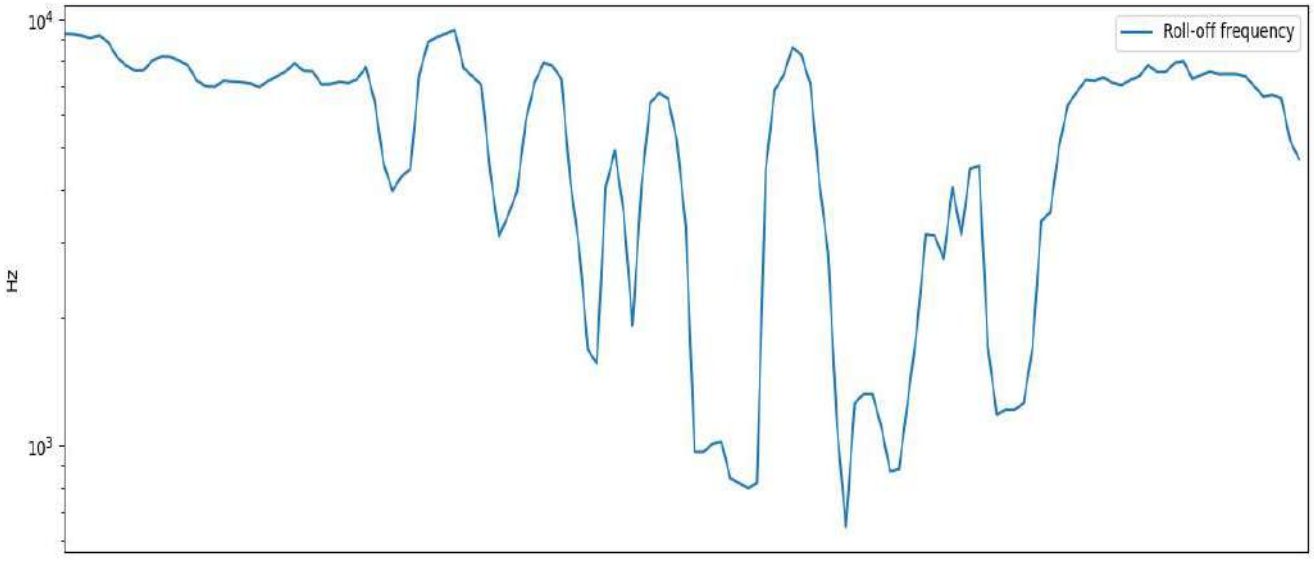


Рисунок 4.15 – Діаграма спектрального спаду природного голосу

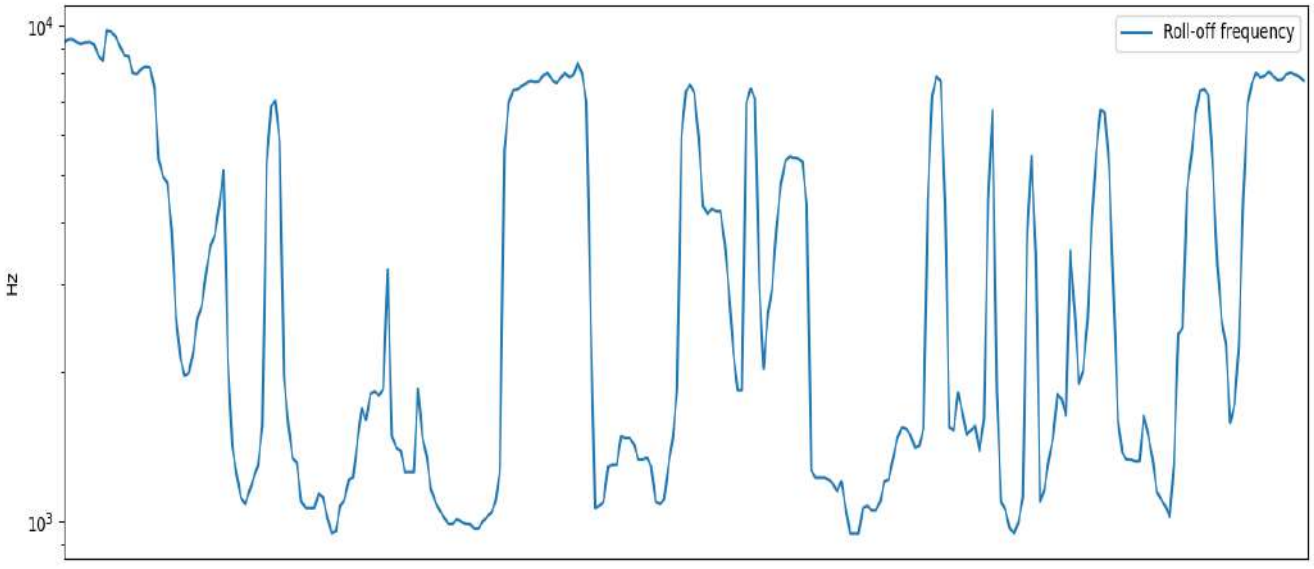


Рисунок 4.15 – Діаграма спектрального спаду штучного голосу

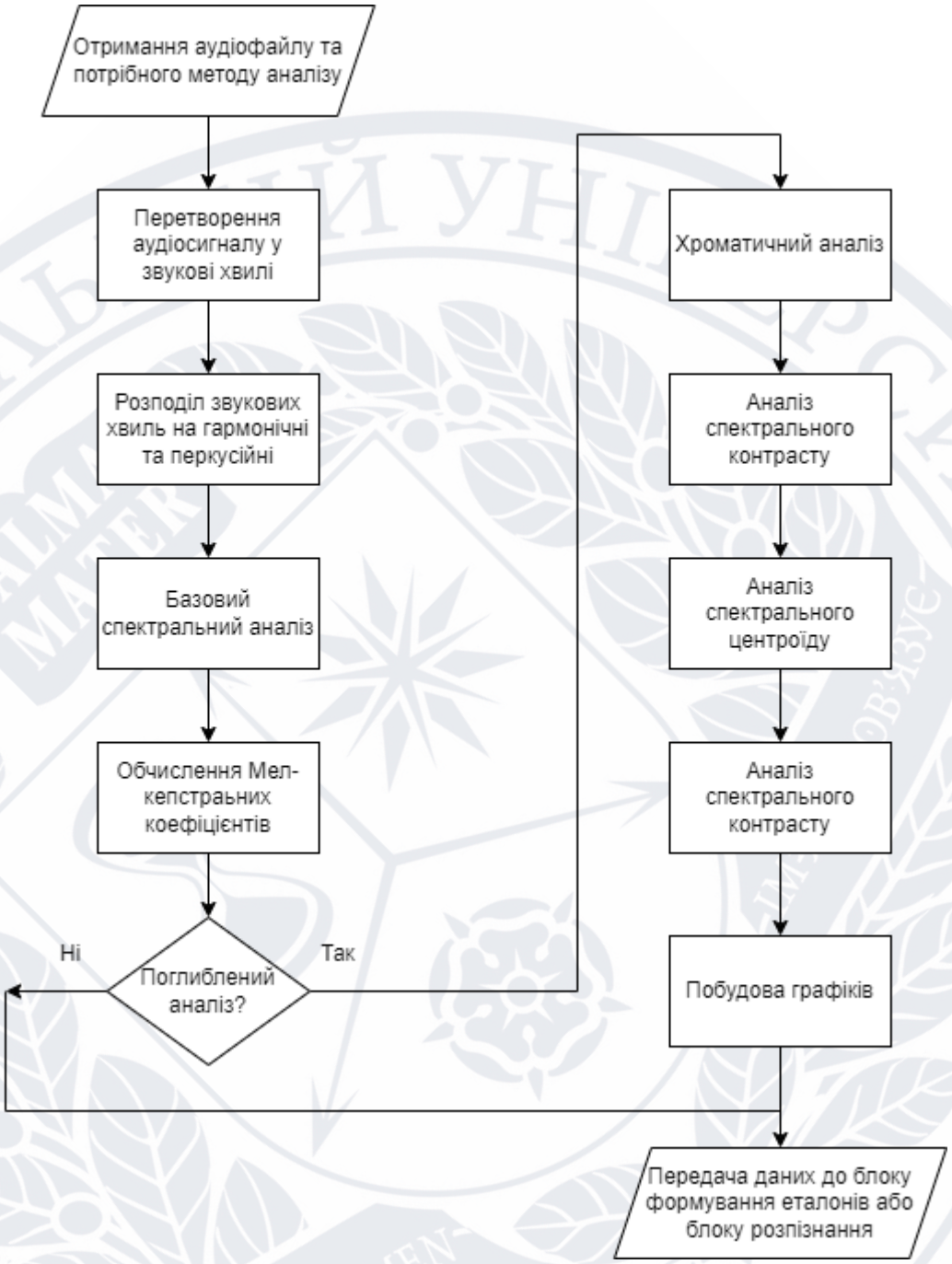


Рисунок 4.17 – Блоксхема покращеного блоку визначення ключових параметрів

Висновок до розділу 4

У цьому розділі було розглянуто розробку додатку. Наведені візуальні відображення результатів виокремлення та обрахунку ключових коефіцієнтів аудіосигналу.

ВИСНОВКИ

У цій кваліфікаційній (магістерській) дипломній роботі було розглянуто актуальність проблеми аналізу аудіосигналу, постановку задачі роботи, а також проведено огляд існуючих програмних продуктів даної тематики, були окреслені їх переваги та недоліки.

Було розглянуто існуючі математичні моделі та методи аналізу аудіосигналу. Було оглянуто стандартну універсальну математичну модель аналізу аудіосигналу за допомогою Мел-кепстральних коефіцієнтів, висвітлено принцип її роботи. Було детально розглянуто основні інструменти аналізу аудіосигналу, та висвітлена специфіка реагування людського вуха за графіком гучності Флетчера-Мунсона а також інструментів аналізу на інструментів аналізу на аудіосигну. Визначено ключові характеристики для генерації штучного голосу, що дає розуміння основних актуальних аспектів аналізу аудіосигналу. Розглянуто специфіку змішаного методу генерації синтетичного голосу, виділені основні тези та кроки його генерації, проведено їх аналіз, та запропоновано методику альтернативну методику глибокого аналізу аудіосигналу, що включає широкий аналіз його спектральних характеристик. Проведено огляд інструментів глибокого спектрального аналізу аудіо даних. Розглянуто галузі їх використання, а також специфіку та принципи їх роботи. Наведено приклади візуалізації їх роботи. Висвітлено технічні складнощі та нюанси, що можуть виникнути в процесі генерації змішаного типу. На основі цього обґрунтовано необхідність створення можливості проведення глибокого спектрального аналізу аудіосигналу лише для конкретних випадків з метою заощадження апаратних потужностей аналізуючої системи, а також збільшення її швидкодії.

Перелічено інструменти, використані для виконання поставленої задачі. Також було створено власну програмну реалізацію запропонованого методу програмного аналізу аудіосигналу.

СПИСОК ЛІТЕРАТУРИ

1. Безпека та злочини, пов`язані з генерацією голосу [Електронний ресурс] – Режим доступу до ресурсу: https://kids.frontiersin.org/articles/10.3389/frym.2022.702664
2. Базовий звуковий аналіз [Електронний ресурс] – Режим доступу до ресурсу: https://dystosvita.org.ua/mod/page/view.php?id=1120
3. Проблеми аналізу звуку різної частотності [Електронний ресурс] – Режим доступу до ресурсу: https://er.nau.edu.ua/handle/NAU/44786
4. Виділення аудіосигналу на фоні шуму з використанням методу сингулярного спектрального аналізу [Електронний ресурс] – Режим доступу до ресурсу: https://eightify.app/uk/summary/technology-and-innovation/advanced-acoustic-cameras-and-software-for-industrial-sound-analysis
5. Програмний застосунок синтезу мовлення Praat [Електронний ресурс] – Режим доступу до ресурсу: https://en.wikipedia.org/wiki/Praat
6. Аналіз звуків мови в Praat [Електронний ресурс] – Режим доступу до ресурсу: http://feltran.kpi.ua/article/view/228388
7. Аналіз звуків мови у програмі Praat [Електронний ресурс] – Режим доступу до ресурсу: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwi0mJOhkP_-AhX7isMKHSIxDuYQFnoECBQQAQ&url=http%3A%2F%2Ffeltran.kpi.ua%2Farticle%2Fdownload%2F228388%2F236387&usg=AOvVaw3r2tkcDs6FRbBMCw2jcf
8. Програмне забезпечення [Електронний ресурс] – Режим доступу до ресурсу: Wavefurfer https://wavesurfer.xyz
9. Переваги та недоліки Wavesurfer [Електронний ресурс] – Режим доступу до ресурсу: https://sourceforge.net/projects/wavesurfer/
10. Створення бібліотек : [Електронний ресурс] – режим доступу до ресурсу: https://metanit.com/sharp/tutorial/3.46.php
11. Іан Соммервіллем Інженерія програмного забезпечення = Software Engineering. — 6-е вид. — М.: «Вильямс», 2002. — С. 642.

12. Джек Грінфілд, Кіт Шорт, Стів Кук, Стюарт Кент, Джон Крупи Software Factories: Assembling Applications with Patterns, Models, Frameworks, and Tools. — М.: «Діалектика», 2006. — С. 592.
13. Min Xu (2004). HMM-based audio keyword generation. У Kiyoharu Aizawa; Yuichi Nakamura; Shin'ichi Satoh. Advances in Multimedia Information Processing – PCM 2004: 5th Pacific Rim Conference on Multimedia
14. Akansu, Ali N.; Haddad, Richard A. (1992), Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets, Boston, MA: Academic Press, ISBN 978-0-12-047141-6
15. Методи аналізу звуку, та їх ключові параметри [Електронний ресурс] – режим доступу до ресурсу: https://virtualorator.com/blog/voice-analysis-understanding-the-report/
16. Пилипенко В. О., Слюсарь І. І., Слюсар В. І. Варіант використання нейронної мережі в системі «Smart Home»
17. Stevens Stanley Smith. A scale for the measurement of the psychological magnitude of pitch / Stanley Smith Stevens, John Volkman & Edwin Newman // Journal of the Acoustical Society of America. – 8 (3). – P. 185–190.
18. Точність аналізу аудіосигналу [Електронний ресурс] – режим доступу до ресурсу: https://callminer.com/faq/how-accurate-is-voice-analysis
19. Acoustic Analysis: New Ideas [Електронний ресурс] – режим доступу до ресурсу: http://asa.scitation.org/journal/publications
20. News and Education/Outreach: [Електронний ресурс] – режим доступу до ресурсу: http://acousticalsociety.org/news/outreach
21. Спектрограми [Електронний ресурс] – Режим доступу до ресурсу: https://musiclab.chromeexperiments.com/Spectrogram/
22. Тембральна унікальність голосу [Електронний ресурс] – Режим доступу до ресурсу: https://ocnt.com.ua/tembralne-zvuchannya-golosu-ta-jogo-fizychni-osoblyvosti/
23. "Overtones and Harmonics". hyperphysics.phy-astr.gsu.edu. Retrieved 2020-10-26.

24. Fineberg, Joshua (2000). "Guide to the Basic Concepts and Techniques of Spectral Music" (PDF). Contemporary Music Review. 19 (2): 81–113.
25. Тональні зміни аудіосигналу [Електронний ресурс] – Режим доступу до ресурсу: https://helpcenter.celemony.com/M5/pdf/melodyneStudio5/en?env=standAlone
26. Бровченко Т. О., Бант І. Н. Фонетика англійської мови. — К.: Рад. шк., 1964. — 270 с.
27. Верба Л. Г. Порівняльна лексикологія англійської та української мов. — Вінниця: Нова книга, 2003. — 160 с.
28. Guido van Rossum, Python Reference Manual, release 2.4.4, 18 October 2006
29. Переваги та недоліки мови Python [Електронний ресурс] – Режим доступу до ресурсу: https://geek.justjoin.it/все-що-ви-маєте-знати-про-python-які-в-нього-н/#Переваги
30. Librosa Library [Електронний ресурс] – Режим доступу до ресурсу: https://librosa.org/doc/latest/index.html

Додаток 2 до наказу
від «31» березня 2023 року
№119/05

ДЕКЛАРАЦІЯ

про дотримання академічної доброчесності

Я, _____
_____

Повністю вказується ПІБ та статус (посада для працівників, освітня (освітньо-наукова) програма – для здобувачів вищої освіти)

що нижче підписалась/підписався, розуміючи та підтримуючи загальновизнані засади справедливості, доброчесності та законності,

ЗОБОВ'ЯЗУЮСЬ:

дотримуватися принципів та правил академічної доброчесності, що визначені законодавством України, локальними нормативними актами Донецького національного університету імені Василя Стуса, положеннями, правилами, умовами, визначеними іншими суб'єктами, та не допускати їх порушення.

ПІДТВЕРДЖУЮ:

що мені відомі положення статті 42 Закону України «Про освіту»;

що у даній роботі не представляла/представляв чиїсь роботи повністю або частково як свої власні. Там, де я скористалася/скористався працею інших, я зробила/зробив відповідні посилання на джерела інформації;

що дана робота не передавалась іншим особам і подається вперше, не порушує авторських та суміжних прав закріплених статтями 21-25 Закону України «Про авторське право та суміжні права», а дані та інформація не отримувались в недозволений спосіб.

УСВІДОМЛЮЮ:

що ця робота може бути перевірена університетом на плагіат або інші порушення академічної доброчесності, в тому числі з використанням спеціалізованих сервісів;

що у разі порушення академічної доброчесності, до мене можуть бути застосовані процедури, передбачені законодавством України та Кодексом академічної доброчесності та корпоративної етики Донецького національного університету імені Василя Стуса, іншими локальними нормативними актами університету, та я можу бути притягнута/притягнутий до академічної відповідальності.

_____ _____
(дата) (підпис)