ISY5001 Intelligent Reasoning Systems

Master of Technology in Intelligent Systems

------------------ PROJECT REPORT ------------------

Unlocking the Gateway to Music: A New Assistant for Music Learning

----

Aug-Oct 2023

Group 16

Team Members

YI YING	A0285683M
ZHANG NI	A0285674M
LIU JINGYI	A0285813X

Abstract

A large audience interested in music education, although the high tuition fees and popularity of piano and guitar lessons pose challenges. The project addresses issues such as evaluation consistency, cost, and noise reduction to enhance the music learning experience. The system's architecture combines Convolutional Neural Networks (CNN) for music analysis, OpenL3 for audio feature extraction, OpenAI for personalized suggestions and Item-KNN for music recommendations. Data preparation involves acquiring and preprocessing training and test datasets, while piano detection and similarity calculations are performed using CNN and OpenL3. The recommendation system utilizes text similarity and Item-KNN to provide personalized piano music suggestions.

Keywords: CNN, OpenL3, OpenAI, Item-KNN, Piano performance scoring, Recommendation system

Table of Contents

1. Introduction 4
1.1 Background 4
1.2 Project Objectives 4
2. Problem Description and Market Analysis 5
2.1 Market Research 5
2.1.1 Expansive Audience 5
2.1.2 Tuition Fee Diversity 5
2.1.3 Prevalent Music Courses 6
2.2 Problem Statement 6
2.3 Competitors Analysis 7
2.3.1 Multidimensional Evaluation 7
2.3.2 Precision 7
2.3.3 Personalization 8
2.3.4 User-Centric Feedback 8
3. System Architecture and Modeling 9
3.1 System Architecture 9
3.1.1 Input Layer 10
3.1.2 Processing Layer 10
3.1.3 Model Layer 11
3.1.4 Data Set 11
3.1.5 Runtime Environment 12
3.2 CNN 12
3.2.1 Convolutional Layer 12
3.2.2 Pooling Layer 13
3.2.3 Fully Connected Layer 13
3.2.4 Activation Function 13
3.3 OPENL3 13
3.4 Item-KNN 14
4. Solution Implementation 16
4.1 Overview Design 16
4.1.1. Multi-dimensional Scoring 16
4.1.2. Precision in Comparison 16
4.1.3. Personalized Comments and Suggestions 16
4.1.4. Music-Based Recommendations System 17
4.2 System GUI Design 17
4.3 System Backend and Model 20
4.3.1 Data Preparation 20
4.3.2 Piano Detection 25
4.3.3 Similarity Calculation 27
4.3.4 Generate Comments 29
4.3.5 Recommendations for piano music 30

4.4 System Evaluation 31
4.4.1 Developer Testing 32
4.4.2 Usability Testing 33
5 Conclusion 36
5.1 Summary 36
5.2 Limitation 36
5.3 Future Plan 36
6 References 38
7 Appendix 39
7.1 Appendix of report: Project Proposal 39
7.2 Appendix of report: Mapped System Functionalities against knowledge, techniques, and skills of modular courses: MR, RS, CGS 43
7.3 Appendix of report: Installation and User Guide 44
7.3.1 Installation 44
7.3.2 User Guide 44
7.4 Appendix of report: 1 or 2 pages individual project report per project member 48
7.4.1 Individual Report – Yi Ying 48
7.4.2 Individual Report – Jingyi Liu 50
7.4.3 Individual Report – Zhang Ni 51

1. Introduction

1.1 Background

In today's fast-paced world, there is a growing trend among adults to prioritize the development of their hobbies, particularly in the realm of music. As they navigate their busy professional lives, many aspire to master their favorite musical instruments with ease and convenience.

Unlike children, who often have the luxury of time for structured music lessons, adults typically find themselves constrained by their schedules. This limitation often leads them to opt for self-guided learning, a flexible approach that enables them to pursue their musical aspirations independently. However, this autonomy comes with its own set of challenges. Self-learners often grapple with uncertainties about the correctness of their playing techniques and struggle to gauge their progress and skill level accurately.

In response to this growing need, there arises a demand for a dedicated application that can provide invaluable support in the journey of adult music learners. Such an app would play a pivotal role in evaluating and tracking their learning progress, offering guidance, and ensuring that they are on the right path to musical mastery.

1.2 Project Objectives

Traditional music assessments typically provide rudimentary scores or ratings. In contrast, our innovative software is designed to offer a more sophisticated evaluation framework, meticulously scoring users' musical performance across multiple dimensions. This translates into a more comprehensive and nuanced understanding of musical prowess, extending well beyond a mere overall score.

Furthermore, our software transcends its role as a mere assessment tool. It transforms into a comprehensive platform that not only evaluates but also furnishes tailored recommendations. These recommendations are meticulously curated based on the identified weaknesses in users' musical performance, facilitating focused and targeted improvement.

2. Problem Description and Market Analysis

2.1 Market Research

Our market research endeavors aimed to gain comprehensive insights into the music education industry in Singapore. This exploration encompassed an assessment of the industry's scope, distinguishing features, and the most popular segments. The following key findings were extracted:

2.1.1 Expansive Audience

Commencing our analysis with the scale of the audience, data from the Singapore Cultural Statistics of 2015 revealed a total of 1,671 music companies operating in Singapore. Notably, among these establishments, 123 brands exclusively offer music-related courses. This substantial presence of music entities underscores the notable popularity and influence of music in Singapore, attracting a broad and diverse audience.

2.1.2 Tuition Fee Diversity

Moving forward, we delve into the intricacies of tuition fees within the music education landscape. The pricing for music instruction in Singapore exhibits a considerable range, spanning from $30 to $200 per month per lesson. While such variations may occasionally create the impression of costly music education, it is imperative to acknowledge the underlying factors. These fees encompass the remuneration of specialized music instructors, essential equipment, and other educational resources that significantly contribute to the total cost. Ultimately, the inherent value of music education transcends the financial aspect, bestowing students with invaluable music skills and profound cultural experiences.

2.1.3 Prevalent Music Courses

Concluding our exploration, we spotlight the most sought-after music courses in the region. Notably, piano lessons reign supreme as the most popular choice, closely followed by guitar lessons. This prevalent trend underscores a substantial proclivity among individuals to acquire proficiency in piano and guitar. These instruments' exceptional musical attributes and expressive capacities captivate a significant portion of the student community, rendering them quintessential instruments in the music education sector.

2.2 Problem Statement

In the realm of music education, learning musical instruments and music theory has always been an exciting yet challenging task. Traditional music learning often involves hiring music teachers, purchasing expensive instruments and materials, and adhering to strict schedules. This has led people to turn to music learning app.

But, with the growth of the market, certain problems have emerged in today's music learning app. One of them is the inconsistency in evaluation. Some apps use the so-called ABCD system to roughly rate songs, but this system may not be accurate enough. Users often wish to obtain more detailed and precise information to better understand the music being played.

Another issue is the cost. Some music recognition apps are priced too high, which may limit their accessibility to some users. We need more reasonably priced options to cater to the needs of different user groups.

Lastly, noise reduction is a significant challenge in the field of music recognition. Some apps are unable to effectively eliminate the impact of surrounding environmental sounds on music recognition. This means that in noisy environments, the accuracy of recognition may decrease, which is clearly a problem that needs to be addressed.

By addressing issues such as evaluation, cost, and noise reduction, we can further drive the development of this field and provide users with a better music recognition experience.

2.3 Competitors Analysis

This section delves into a comprehensive competitive analysis in the realm of music, with a specific focus on our new music assessment system's four primary dimensions: multidimensional evaluation, precision, personalization, and feedback, all of which represent the central tenets of our innovative approach.

2.3.1 Multidimensional Evaluation

Our system's cornerstone feature, multidimensional evaluation, warrants primary attention. In stark contrast to traditional music assessments, which often furnish a singular overall score, our pioneering system employs artificial intelligence to assess music across a multitude of dimensions. This multipronged evaluation approach yields a far more exhaustive and insightful score, enabling an in-depth exploration of musical performance and affording users exceptionally valuable assessments.

2.3.2 Precision

Precision stands as another pivotal facet of our system. It allows for the meticulous comparison of a user's music with established industry standards. Our system excels in providing users with a precise understanding of how closely their musical performance aligns with these standards, serving as a yardstick for improvement and refinement.

2.3.3 Personalization

Our commitment to personalization is unwavering. Our system is not merely an assessment tool but also a provider of tailored recommendations. These recommendations are rooted in individual weaknesses identified during the assessment process. Users receive guidance on specific areas of improvement and corresponding practice materials, ensuring that their journey towards musical mastery remains highly personalized.

2.3.4 User-Centric Feedback

The user's voice is paramount in our system's evolution. We accord the utmost significance to user feedback, fostering a dynamic feedback loop. This mechanism facilitates continuous refinement and enhancement based on user insights. The feedback collected directly influences the evolution of our artificial intelligence systems, perpetually aligning the technology with user expectations and requirements.

In sum, our innovative music assessment system signifies a paradigm shift, delivering multidimensional evaluations, precision in assessments, personalized recommendations, and an unwavering commitment to user-centric enhancements through continuous feedback. These elements coalesce to empower music enthusiasts, enabling them to comprehend and elevate their musical prowess with unmatched precision and depth of understanding.

3. System Architecture and Modeling

3.1 System Architecture

The system leverages advanced technology and machine learning models to provide a user-centric learning experience, emphasizing multi-dimensional evaluation, precision in comparisons, personalized recommendations. The architecture of the system encompasses several key components, each contributing to the overall functionality and user experience.

The framework initially processes the user-provided MP3 audio files, performing format conversion and feature extraction. These extracted features are then input into a CNN network for piano music detection, followed by similarity calculations between the input audio and a standard music library using the OpenL3 model. The results of this similarity analysis are used to generate personalized user evaluations and recommendations, tailored to the specific piano compositions, enhancing the user experience and learning process.

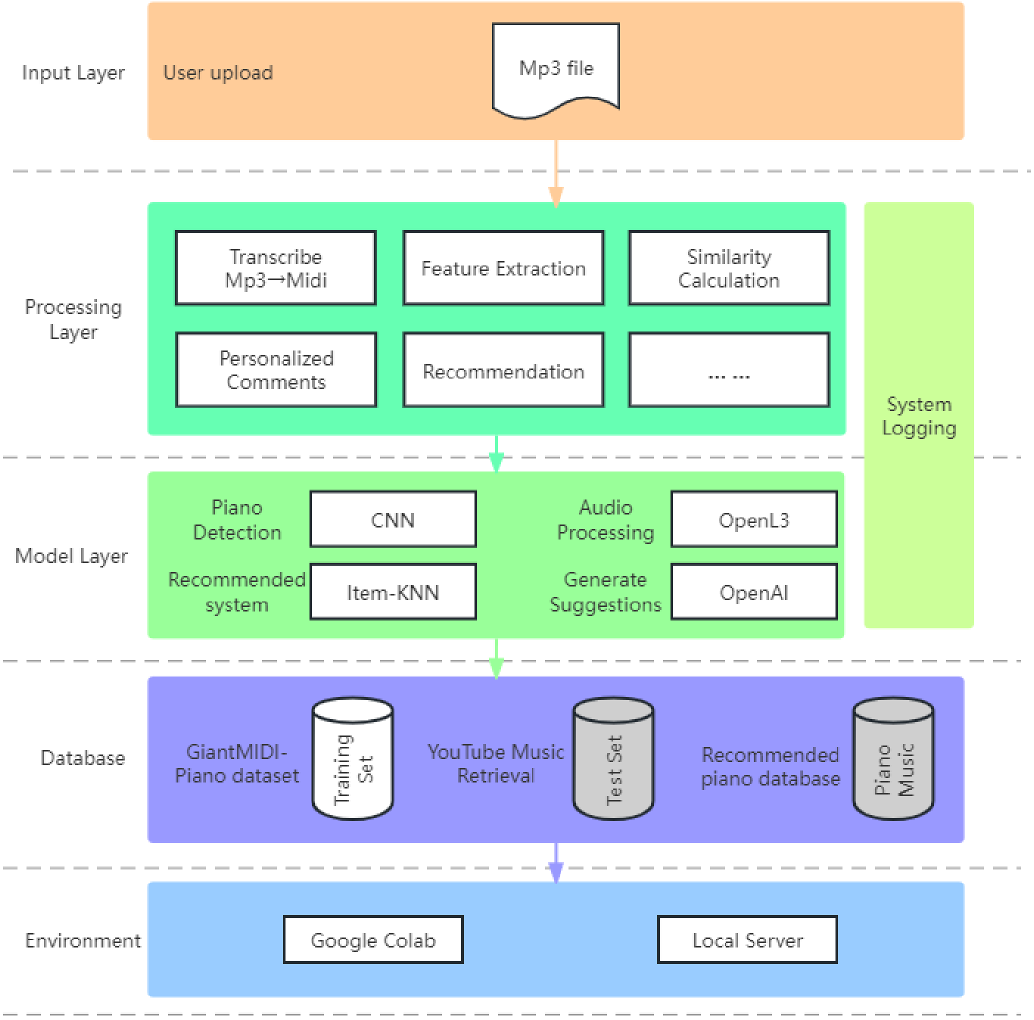


Fig. 1. System Architecture

3.1.1 Input Layer

At the input layer, users have the ability to upload their piano performances in MP3 format. This layer serves as the initial point of interaction with the system.

3.1.2 Processing Layer

The processing layer is the core of the system, where several essential operations take place:

1. MP3 to MIDI Conversion: Uploaded MP3 files are converted into MIDI format to

extract musical data for analysis.

2. Audio Feature Extraction: The system extracts crucial audio features from MIDI files, capturing aspects such as pitch, rhythm, and timbre.
3. Similarity Calculation: Music similarity is calculated, enabling accurate comparisons between user-generated performances and standard compositions.
4. Personalized Comments and Recommendations: Based on extracted features and similarity calculations, the system generates individualized comments, ratings, and recommendations for each user, thereby enhancing the learning experience.
5. System Logging: The system maintains logs of activities and user interactions for transparency and performance monitoring.

3.1.3 Model Layer

The model layer integrates various sophisticated models to perform diverse tasks:

CNN (Convolutional Neural Network): This model is employed to detect whether uploaded audio files are piano compositions.

OpenL3: OpenL3's pre-processing model is used for audio analysis and similarity calculations, assessing aspects like tonality, rhythm, and other musical characteristics.

OpenAI API: The OpenAI API interface is used for generating personalized comments, ratings, and instructional feedback tailored to specific users.

Item-KNN (Item-based k-Nearest Neighbors): This model recommends related piano compositions based on user preferences and musical style.

3.1.4 Data Set

The system relies on a robust data foundation, which includes the following:

Training Dataset: The training dataset is sourced from the open-source GiantMIDI-

Piano database, providing a comprehensive dataset for model training.

Test Dataset: Additional data is collected through Python web scraping, retrieving piano music data from YouTube. This dataset serves as the basis for testing and evaluating the system's performance.

3.1.5 Runtime Environment

The system operates in a dual runtime environment:
a. Google Colab: Google Colab is utilized for accessibility and scalability, allowing users to interact with the system through a web-based interface.
b. Local Server: A local server complements the online platform, ensuring system stability and performance, particularly for computationally intensive tasks.

3.2 CNN

A Convolutional Neural Network (CNN) is a specialized deep learning model for processing visual data, like images. It's excellent at recognizing patterns and features in images and is commonly used in tasks like image recognition and object detection. CNNs consist of layers that learn to extract and understand visual information. They are widely used in computer vision applications. There are some key components of CNN: Convolutional Layer, Pooling Layer, Fully Connected Layer, Activation Function.

3.2.1 Convolutional Layer

The convolutional layer is the core component of CNN. It includes a set of convolutional kernels (also known as filters) used to perform convolution operations on input data. This operation efficiently captures local features, such as edges and textures in images. The convolutional kernels slide over the entire input, generating an output known as a feature map.

3.2.2 Pooling Layer

The pooling layer is used to reduce the spatial dimensions of feature maps while retaining the most important information. The most common pooling operation is max pooling, where the maximum value within each pooling window is retained, reducing the data's dimensionality. This helps reduce computational burden, mitigate overfitting, and improve the model's invariance.

3.2.3 Fully Connected Layer

Fully connected layers are typically positioned at the top of the CNN and receive features from the preceding layers to generate the final output. This layer is similar to a traditional feedforward neural network, where each neuron is connected to all neurons in the previous layer.

3.2.4 Activation Function

Activation functions introduce non-linearity, allowing the model to capture more complex patterns. Common activation functions include ReLU (Rectified Linear Unit), Sigmoid, and Tanh.

The core idea of CNN is local perception and weight sharing, enabling it to learn features on a large scale and exhibit exceptional performance. This makes it the model of choice for many modern deep learning tasks.

3.3 OPENL3

OpenL3 is an open-source deep learning model used to extract high-level audio embeddings from audio data. These embeddings represent semantic information in the audio, making OpenL3 a valuable tool for diverse audio analysis tasks. Its open-source nature, support for various audio formats, and applications in music recommendation, emotion analysis, and content analysis make it a versatile and accessible resource for audio analysis. Researchers and developers can easily integrate OpenL3 into their workflows for tasks like classification and feature extraction.

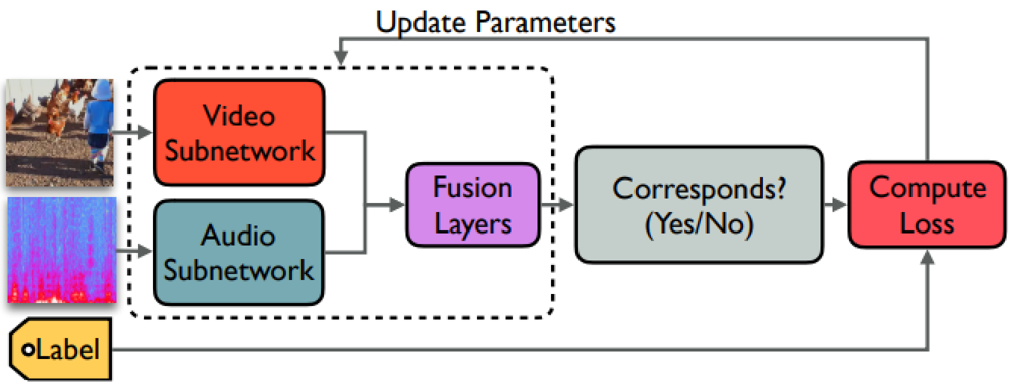


Fig. 2. High-level architecture of L^3–Net [7,8].

3.4 Item-KNN

The item-KNN method is performed to provide personalized recommendations by finding similarities between items based on user interactions and preferences.

The basic idea behind item-KNN is to analyze the historical data of user-item and identify items that are similar to each other. This similarity is calculated using a distance metric called standard Euclidean distance.

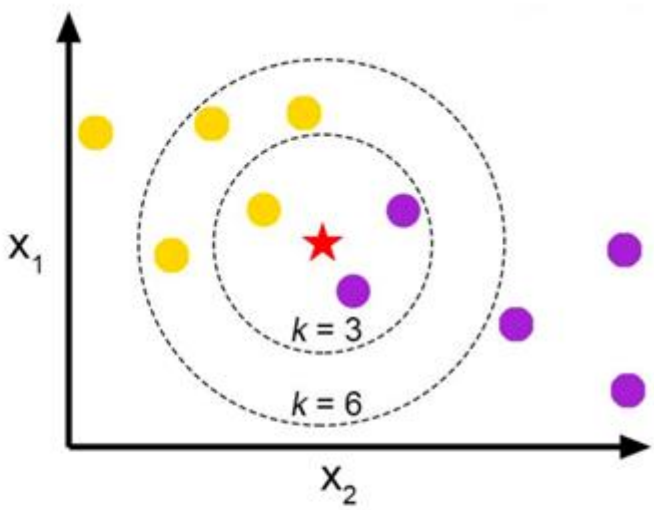


Fig. 3. KNN method

The algorithm then sort all points in order of distance from low to high, and pick up 5 points with the smallest distance to be the final recommended results.

4. Solution Implementation

4.1 Overview Design

In response to market requirements, our system has been meticulously crafted, adhering to fundamental principles encompassing multi-dimensional scoring and precision in comparison. These principles serve as the bedrock of our approach, ensuring the provision of an intricate and user-centric solution for audio analysis and music recommendations.

4.1.1. Multi-dimensional Scoring

Our system adopts a multi-dimensional rating approach to assess various aspects from multiple dimensions. This approach provides a comprehensive evaluation, ensuring a holistic understanding of the audio content.

4.1.2. Precision in Comparison

We are committed to precision when comparing user-generated audio with standard music. Our advanced machine learning and signal processing techniques enable us to unlock insights into sound quality and patterns, facilitating precise and insightful comparisons.

4.1.3. Personalized Comments and Suggestions

To enhance user engagement and optimize the learning experience, our system intelligently crafts personalized suggestions, driven by user performance metrics and individualized learning requirements. These personalized suggestions employ individual user data to tailor content in an adaptive manner.

4.1.4. Music-Based Recommendations System

Our system focuses on music-based recommendations, suggesting related piano compositions based on the user's current music selection. This encourages users to practice and explore similar piano pieces, creating an engaging and dynamic learning experience.

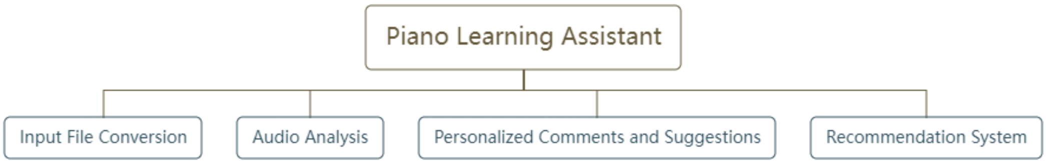


Fig. 4. System overview design.

4.2 System GUI Design

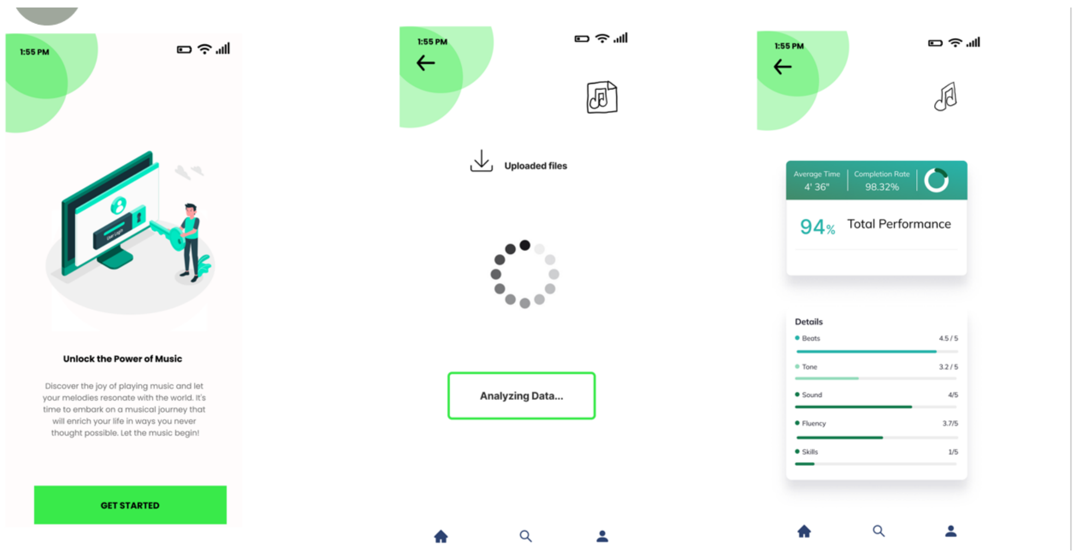


Fig. 5. Use case demo part1.

Our application has a user-friendly interface designed to provide a seamless experience.

1. Upload File Interface (Left Panel):

The left panel of our interface is specifically designed for uploading files. Here, users can easily upload MP3 files by clicking the "Upload" button. This step is the initial interaction point for users to submit audio files for analysis.

2. File analysis interface (middle panel):

After successful upload, the application will seamlessly transition to the file analysis interface. In this section, users can expect the following:

Real time progress update: Users will be able to track the progress of file analysis, ensure transparency, and be notified at any time throughout the entire process.

Detailed analysis: This application will conduct a comprehensive analysis of uploaded MP3 files, including melody, harmony, rhythm, and other related music components.

Visualization: Users will obtain a visual representation of audio file features, enabling them to better understand the analysis results.

3. Scoring interface (right panel):

The right panel of the interface is dedicated to displaying the final score of the uploaded MP3 file. When our advanced model completes the analysis, users can expect the following:

a) Comprehensive evaluation: This application will provide users with detailed ratings to evaluate music quality, complexity, and other related standards.
b) Interactive feedback: Users will have the opportunity to explore their music from a new perspective and gain a deeper understanding of areas they may not have noticed before.
c) Download and Share: In this section, users will be able to download their analysis reports and share them with others, thereby promoting collaboration and discussion.

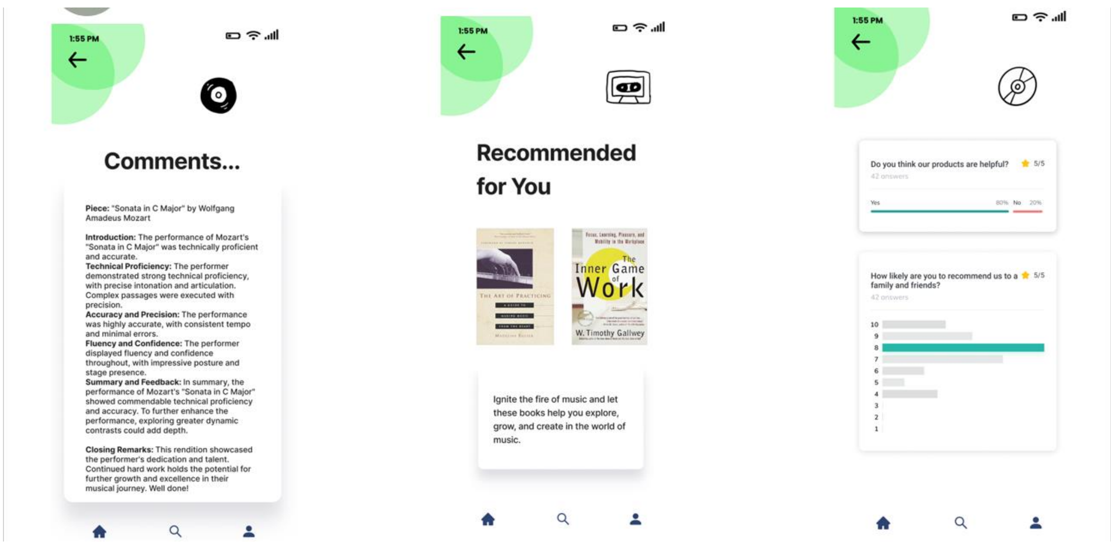


Fig. 6. Use case demo part2.

4. Evaluation interface (left panel):

Feedback and feedback: Users will find detailed feedback and feedback related to their performance. Whether it's musicians, students, or aspiring artists, our platform provides insightful comments on their works, pointing out areas for improvement.

Personalized suggestions: Our evaluation system aims to provide personalized suggestions to help users improve their skills to a new level. Users can use these recommendations to fine-tune their performance and achieve their music goals.

5. Recommendation interface (middle panel):

The middle panel of the interface is specifically designed for recommendation, where users can access carefully selected educational resources. Here are their findings:

a) Tailored book recommendations: Based on users' unique music situation and skill level, our application recommends relevant books, scores, and educational materials to assist users in their practical and learning journey.
b) Skill enhancement: The recommended resources have been carefully selected to supplement users' music journey, providing them with the tools they need to improve their skills and explore new perspectives.

6. Feedback interface (right panel):

The right panel is specifically designed for feedback interface, creating an open communication channel between users and our development team. In this section, users can:

a) Provide feedback: Encourage users to share their thoughts, suggestions, and experiences with our application. We highly value user feedback and strive to continuously improve based on it.
b) Provide suggestions: If users have specific ideas or suggestions for enhancing the functionality of the application, they can express them here at any time. User insights contribute to the sustained development and enhancement of the platform.
c) Collaborative development: We believe in jointly creating a platform that truly meets user needs. Through dialogue with our community, our goal is to establish a dynamic and responsive ecosystem that promotes creativity and learning.

4.3 System Backend and Model

4.3.1 Data Preparation

1. Training datasets

The GiantMIDI-Piano dataset [9] is a vast collection of high-quality MIDI files primarily focused on piano music. It offers a diverse range of piano pieces, including metadata and annotations. This dataset serves as a valuable resource for training and evaluating machine learning models in music-related tasks, such as transcription, chord recognition, and melody extraction. Researchers and developers can benefit from its size and diversity for various music applications.

Dataset	Composers	pieces	Hours	Type
piano-midi.de	26	571	36.7	Seq.
Classicalarchives	133	856	46.3	Seq.
Kunstderfuge	598	-	-	Seq.
MAESTRO	62	529	84.3	Perf.
MAPS	-	270	18.6	Perf.
GiantMIDI-Piano	**2,786**	**10,854**	**1,237**	**Live**

Table 1: Piano MIDI datasets

2. Test Datasets

we have designed the model to create our **test dataset**.

Data Collection and Test Dataset Creation Process:

Our approach to generating a test dataset for our model is a multi-step process, meticulously designed to ensure the acquisition of comprehensive and relevant data. Here's an overview of each step:

a. IMSLP Data Retrieval:

To kickstart the dataset creation process, we initiate by downloading HTML pages from the IMSLP (International Music Library Project) website. These pages encompass a repository of renowned composers' information and compositions. Our approach comprises the following steps: Initial Page Download: We begin by downloading the page that contains a list of all composers.

Composer Name Extraction: After obtaining the composer's list, we extract the names of the composers, which will serve as a key reference point for further data retrieval.

Composer-Specific Page Downloads: With the list of composers in hand, we proceed to download individual HTML pages for each composer to access more detailed information.

b. Wikipedia Biography Retrieval:

In our quest for comprehensive composer data, we access the composer's Wikipedia page to seek detailed biographical information. This step unfolds as follows:Wikipedia Page Download: We download the Wikipedia pages dedicated to each composer, aiming to capture valuable insights into their life, work, and contributions to the world of music.

c. Metadata Compilation:

With the data from IMSLP and Wikipedia in our possession, we embark on the creation of a metadata CSV file. This file acts as a repository for information related to composers, their musical compositions, nationality, birth, and death details. The CSV file is populated using data collected from previously downloaded HTML and Wikipedia pages.

d. YouTube Music Retrieval:

As a crucial part of our dataset creation process, we delve into the realm of music by extracting information from YouTube. Our approach encompasses the following steps:

Video Title and ID Extraction: Based on composer and music names, we extract relevant information from the CSV file generated in the previous steps. This information is used to search for specific music works on YouTube.

Music Download: We proceed to download these music works from YouTube, ensuring that we have access to high-quality audio content.

MP3 Conversion: To standardize the format, we convert the downloaded music files to MP3 format, making them readily usable for subsequent stages of our model development.

3. Piano Transcription with Piano_Transcription_Inference:

To transcribe the MP3 music files into MIDI format, we utilize the powerful Piano_Transcription_Inference toolbox. This toolbox streamlines the transcription process, making it accessible and efficient. It's important to note that the toolbox can be effortlessly installed, and after installation, it enables us to perform the following tasks: Transcription of Piano Recordings: Using this toolbox, we can seamlessly convert piano recordings into MIDI files, ensuring that the musical essence is captured in a digital format.

And we use [Essentia] for graph display. The big strength of Essentia is its extensive collection of optimized and tested algorithms for audio processing and analysis, all conveniently available within the same library.

We first imported the required libraries, including Essentia and Matplotlib, as well as some required files. Defined SEGMENT_ LENGTH, used to specify the length of audio segments in seconds. Use Essentia's EasyLoader class to load audio data from a specified audio file. Using essentia Pool() creates a feature pool pool for storing extracted audio features. In the following code snippets, various feature extractors from Essentia were used, including RMS (root mean square), HFC (high-frequency content), etc., to extract audio features. A window function was created using es.Windowing, and then audio frames were iterated through es.FrameGenerator to extract features for each frame. For each frame, perform the following actions: Calculate RMS: Calculate the root mean square of the audio frame using the RMS function. Calculate Spectrum: Calculate the spectrum of audio frames using the spectrum function. Calculate RMS Spectrum: Similarly, calculate the root mean square of the spectrum. Calculate HFC: Calculate the high-frequency content of audio frames. Finally, visualization was performed using the Matplotlib library to draw and visualize multiple graphics, including audio waveforms. Signal envelope. Visualization of RMS features. Visualization of Spectrum RMS features. Visualization of High frequency content features. The effect is as follows:

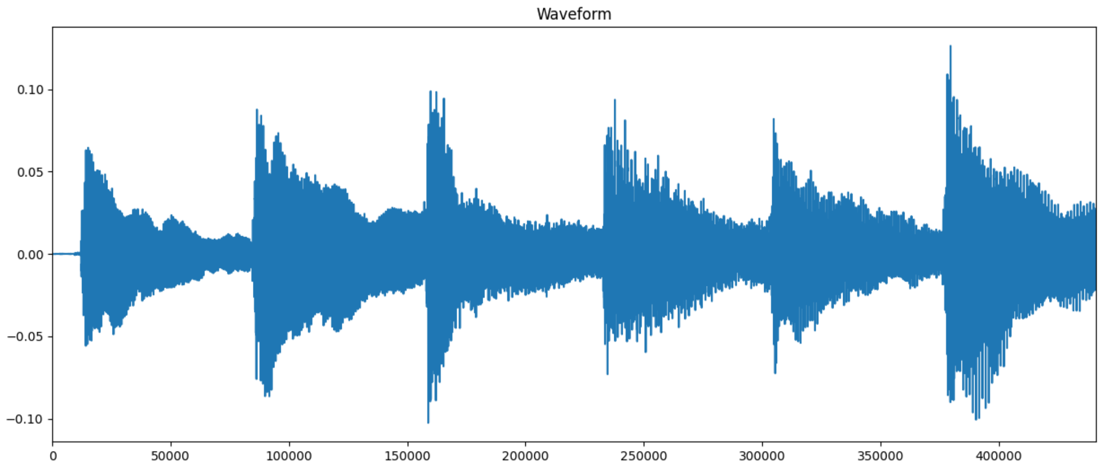


Fig. 7. Waveform graph

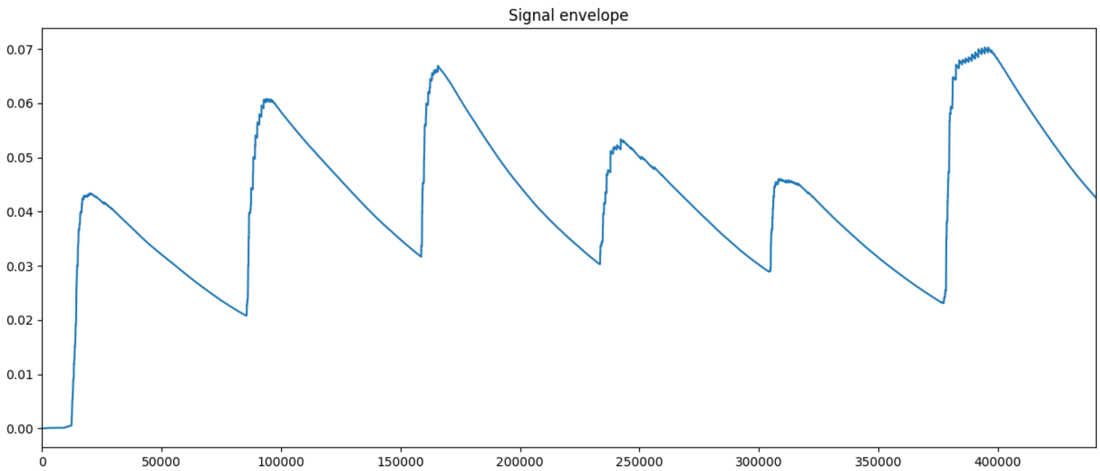


Fig.8. Signal envelope graph

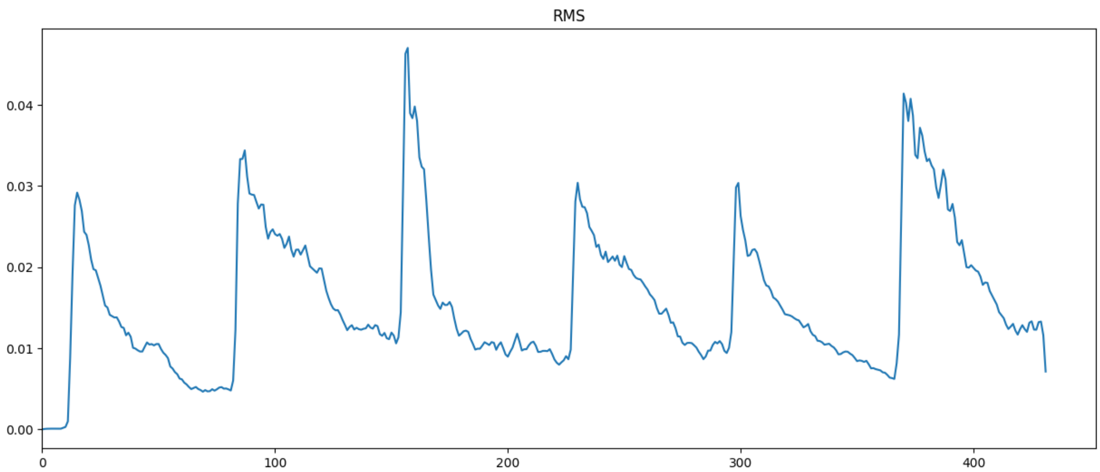


Fig.9. RMS graph

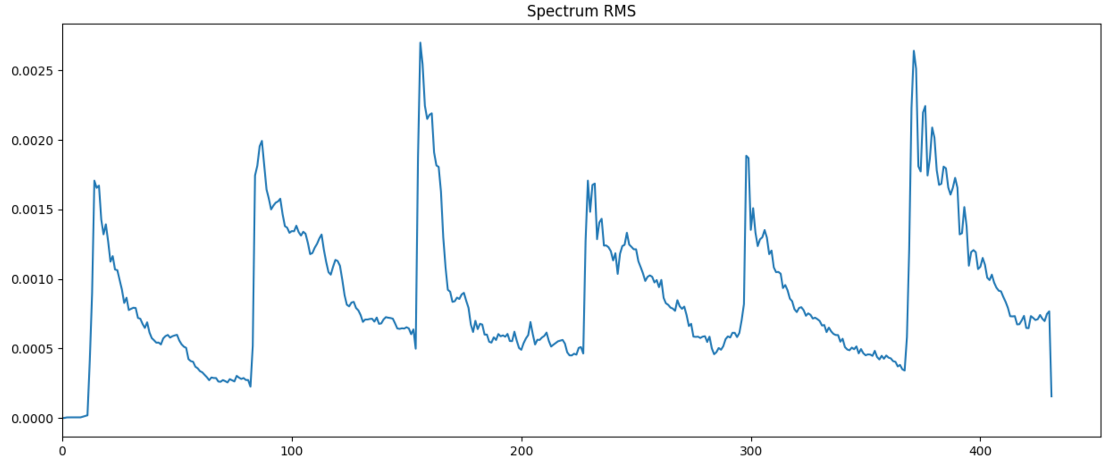


Fig.10. Spectrum RMS

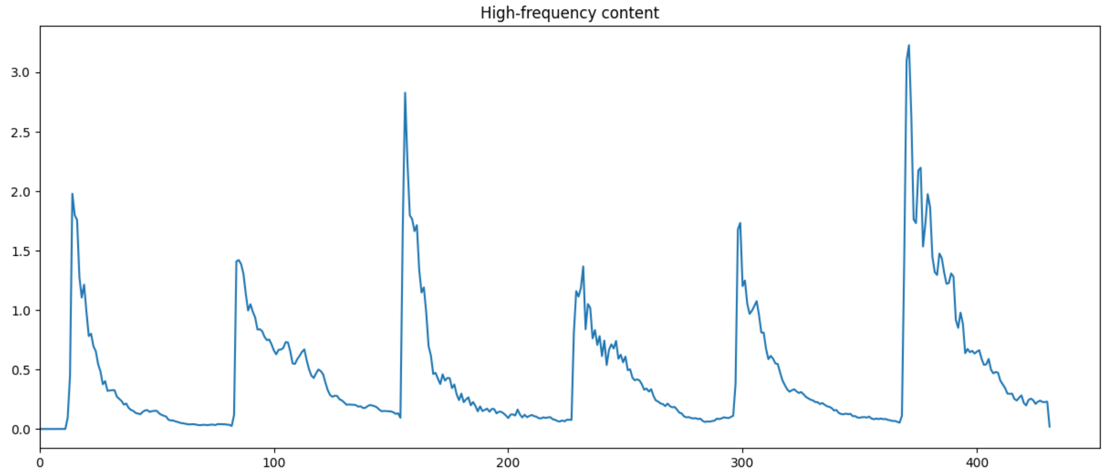


Fig.11. High-frequency content graph

Afterwards, we perform pitch visualization. Install the Essentia library, load audio files, extract pitch curves, calculate the time position of pitch curves, draw pitch curves and pitch confidence. The result is as followed.

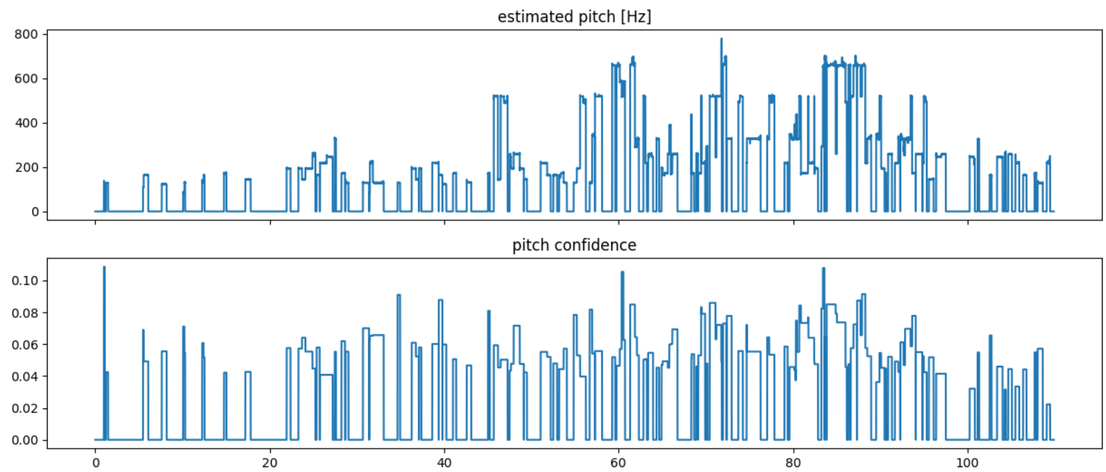


Fig.12. Pitch graph

This comprehensive process ensures that our model is equipped with a diverse and rich test dataset, facilitating its development and refinement. The dataset encompasses information from IMSLP, Wikipedia, and YouTube, harmoniously combined to support the training and evaluation of our model. The use of Piano_Transcription_Inference further enhances our capabilities, enabling us to transcribe audio recordings into a format that's instrumental for our research and development goals.

4. Recommendation data collection and preprocessing

We used methods such as requests and beautifulsoup to crawl 3504 piano music in RenRen Piano Music and save them in CSV file. Their attributes are: music name, author name and music type. Technically, we use python to crawl the HTML file of the web page, and then interpret it and find the required resources.

4.3.2 Piano Detection

Since our app is for scoring pianos, we need to test the piano part. In this project, we used CNN to train the model for piano partial detection.

We have designed the following model:

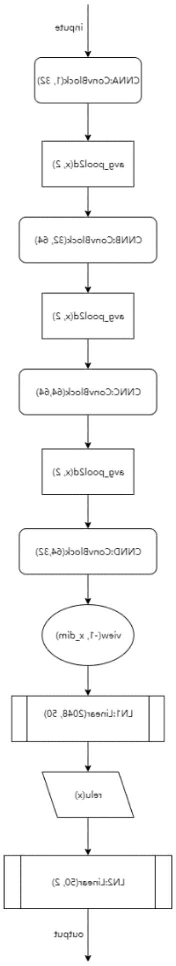


Fig. 13. CNN model structure

The data x starts at the first convolutional block, then undergoes downsampling through pooling layers, and continues through the subsequent convolutional blocks until the final one. It is then flattened into a one-dimensional tensor using the view operation to be input into the fully connected layers.

The followings are the detail :

x = self.CNNA(x): Process through the first convolutional block.

x = avg_pool2d(x, 2): Average pooling layer to downsample the feature maps with a window size of 2.

x = self.CNNB(x): Process through the second convolutional block.

x = avg_pool2d(x, 2): Another round of downsampling.

x = self.CNNC(x): Process through the third convolutional block.

x = avg_pool2d(x, 2): Further downsampling.

x = self.CNND(x): Process through the fourth convolutional block.

x = avg_pool2d(x, 2): Further downsampling.

x_dim = x.shape[1] * x.shape[2] * x.shape[3]: Calculate the input dimension after flattening.

x = x.view(-1, x_dim): Flatten the tensor to one dimension.

x = self.LN1(x): Forward pass through the first fully connected layer involving a linear transformation.

x = relu(x): Apply the ReLU activation function.

x = self.LN2(x): Forward pass through the second fully connected layer, yielding the final prediction output.

We will input audio information, and the model will output judgment.

4.3.3 Similarity Calculation

a. Traditional cosine similarity result calculation

traditional Cosine similarity algorithm: The cosine value between the angles between two vectors in a vector space is used to measure the difference between two individuals. When the cosine value approaches 1 and the angle approaches 0, it indicates that the more similar the two vectors are, the closer the cosine value approaches 0 and the angle approaches 90 degrees, indicating that the two vectors are less similar. The calculation formula is as follows:

$$\cos\_\mathrm{sim} = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| \cdot |\vec{b}|}$$

Firstly, we designed a function called align, which is mainly used to perform music alignment operations. The function first specifies the path to a CSV file csv_ path, this file is used to contain information about music alignment. We have defined align_ tools_ dir variable, which specifies a directory for storing music alignment tools. Two more directories were created, namely_tmp and aligned_ results, used to store temporary files and alignment results. Afterwards, by reading the content of the CSV file, the function stores the file content in the lines list, and traverses the lines list to process each line. For each row, the function extracts the including piece_ name, gt_ name, maestro_ name and giantmidi_ information including name. The function executes a series of commands, including copying MIDI files, aligning music, and so on. Then, it uses the cp command to copy different MIDI files to the directory of the music alignment tool align_tools_dir. Call the music alignment tool MIDIToMIDIAlign.sh again to perform music alignment, targeting gt_name and maestro_ name, and gt_ name and giantmidi_ name aligns with music. Finally, it copies the alignment results back to aligned_results directory.

Then, we have designed a function that uses hard coded CSV_ path variable specifies the path to a CSV file. This CSV file contains music related data. This file is used to provide information on music alignment. The function reads the content of a CSV file and analyzes the data line by line. For each row of data, the function extracts relevant information from it, including time information, pitch information, etc. The function calculates similarity scores for music data, which involve similarities in aspects such as beat score, volume score, and pitch score. The similarity score calculation is completed by calculating cosine similarity. Cosine similarity is a method of measuring the similarity between vectors. Finally, the function prints out the similarity scores of beat, volume, and pitch.

b. Similarity calculation based on OpenL3 pre-trained model

Audio similarity computation is a common task in music information retrieval, content recommendation, and many other audio-related applications. OpenL3 is a powerful tool that leverages deep learning to extract meaningful audio features. In this report, we focus on computing audio similarity based on these features.

OpenL3 Feature Extraction: We load the OpenL3 pre-trained model and use it to extract audio features from the input audio files. The OpenL3 model provides options for specifying the content type and input representation. In this report, we use the default settings for simplicity.

Similarity Calculation: OpenL3 (Open source Deep Learning for Audio Analysis) can be used to easily extract audio features, and then use these features to calculate audio similarity.

c. Results:

The computed audio similarity score represents the degree of similarity between the two audio files. A higher similarity score indicates greater similarity, while a lower score suggests dissimilarity. The results can be used in various applications, such as music recommendation systems, content matching, or audio search.

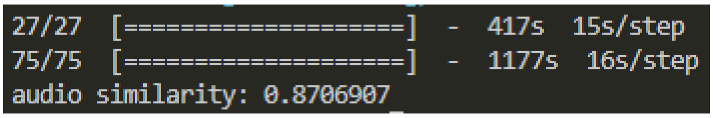


Fig.14. Example of model calculation results

4.3.4 Generate Comments

The system will automatically rate it and generate reviews. Reviews include the quality of the music, sentiment analysis and technical points. In addition, the system will provide personalized suggestions based on ratings and features to help users improve and enrich their musical works. The core of using the OpenAI API is sending HTTP POST requests to the API endpoint. This can be done programmatically using Python. After sending an API request, OpenAI responds with the generated content.

4.3.5 Recommendations for piano music

The utilization of text similarity to build a recommendation system, which involves comparing customer input with a database, is a widely adopted approach for delivering personalized recommendations to users. The main advantages of utilizing text similarity recommendation systems is their ability to provide highly relevant recommendations based on a user's choice. By analyzing the similarities between various piano music and user's preference, the system can suggest piano music that are likely to be appealing to individual users.

Our system algorithm utilizes the TfidfVectorizer to convert the piano music database and customer's input into numerical vector representations, where each vector represents the presence or absence of specific words or terms. Then, Item-KNN method is performed between the user-vector and database to find 5 piano music that is closest to the user preference. To measure the distance between the user-vector and each point in the database, the model use the standard Euclidean distance.

In our piano music recommendation system, we faced a significant challenge. How to deal with chinese charactors in the piano music's name properly. Chinese is not like English, has a natural division of spaces between words. Therefore, we first need to perform word segmentation processing and then convert it into a string format. The famous Chinese word segmentation library jieba is used for this word segmentation.

The second challenge is that, we found single words were not represented in numerical vector. After searching some references, we noticed that the default value of the parameter "token_pattern" only matches words with a length ≥ 2. Generally speaking, words with a length of 1 are usually insignificant in English, but in Chinese, there may be some important single words, so we modified the parameter.

4.4 System Evaluation

We use Essentia to show the similarity.

we plot using a similarity score matrix to draw the graph. The following are the display results with high similarity.

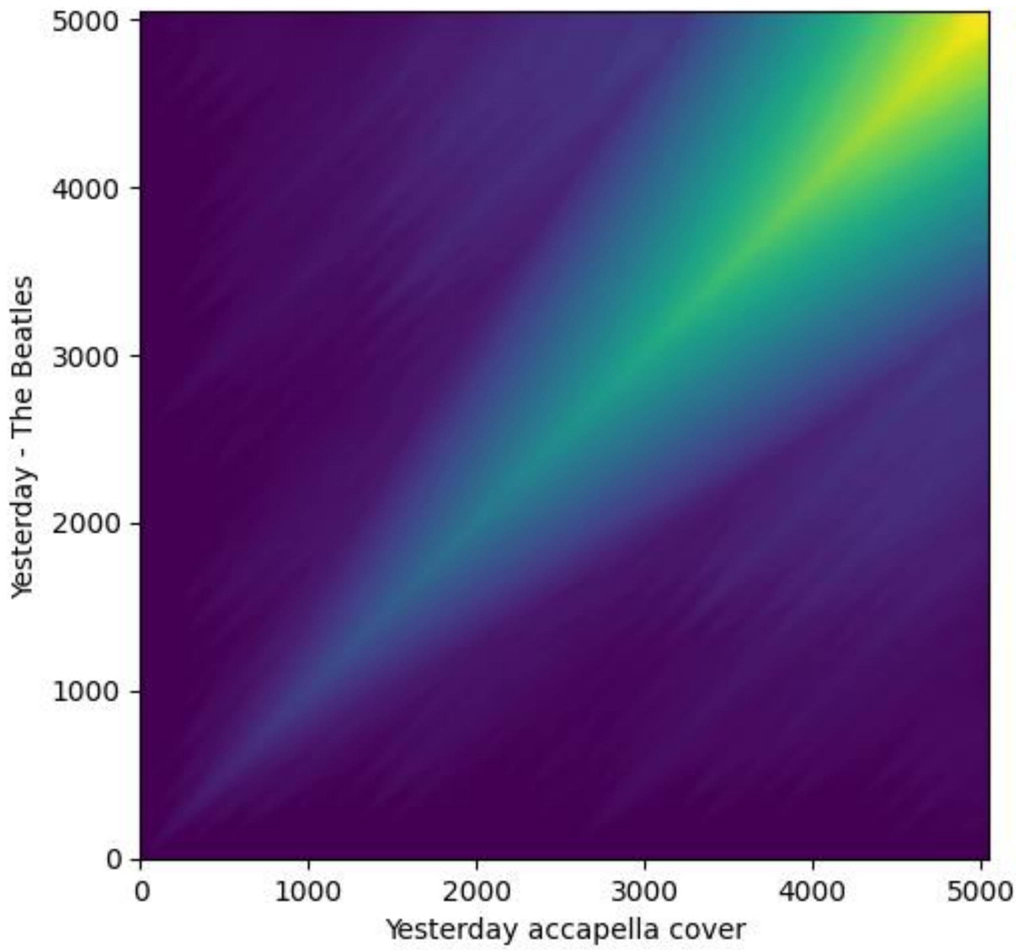


Fig. 15. high similarity graph

The following is a drawing with low similarity.

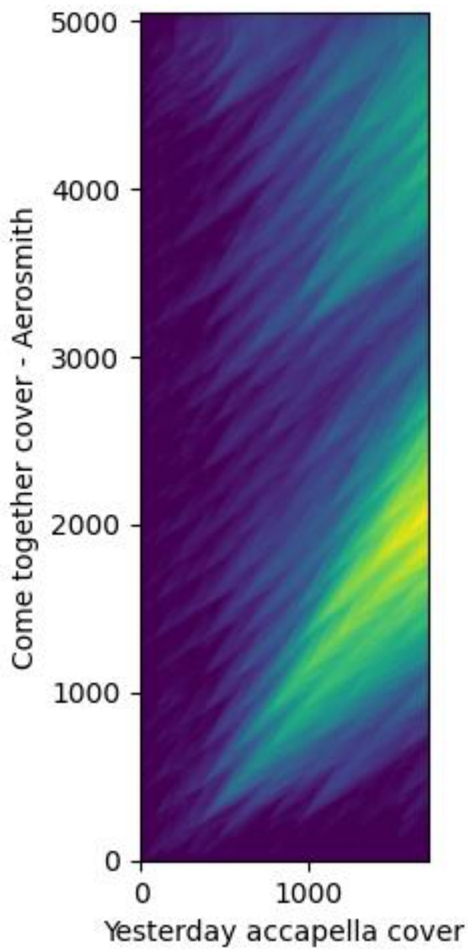


Fig. 16. low similarity graph

4.4.1 Developer Testing

Developer testing is a critical phase in our system development. We use various testing methods to ensure the system's reliability and functionality. Here's an overview of key developer testing metrics:

Testing Method	Metrics	Results
Unit Testing	Percentage of Passed Tests	100%
Performance Testing	Average Response Time (ms)	42.5
Security Testing	Vulnerabilities Identified	12
Code Review	Number of Issues Identified	35

Table 2: Developer Testing

4.4.2 Usability Testing

According to the overview method used in getting the SUS score, we used numbers from 1 to 5 to let users rank every question based on their level of agreement. Moreover, 1 means strongly disagree, 5 means strongly agree. Besides, we subtracted 1 from the score for each of the odd-numbered questions. While, for each of the even-numbered questions, subtract their value from 5. Then, we added these new values to get the total score and multiplied it by 2.5. This method helped us score out of 100, which is very direct. (Thomas, 2022).

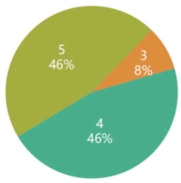
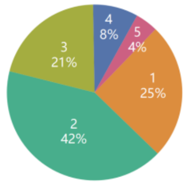
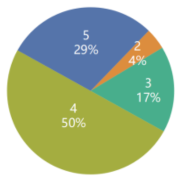
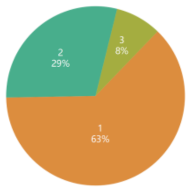
For odd questions, we did the following operation:

(score-1)*2.5

For even questions, we did the following operation:

(5-score)*2.5

After calculation, the higher the result is, the better the comment is.

No.	Question	Average Mark	Pie Chart
1	I think that I would like to use the system.	(score-1)*2.5=8.44	
2	I found the system unnecessarily complex.	(5-score)*2.5=6.88	
3	I thought this system was easy to use.	(score-1)*2.5=7.60	
4	I think that I would need the support of a technical person to be able to use this system.	(5-score)*2.5= 8.85	

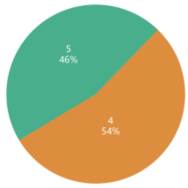
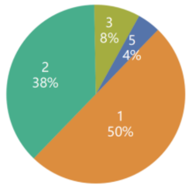
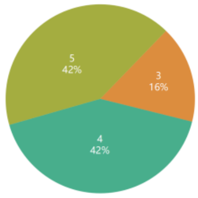
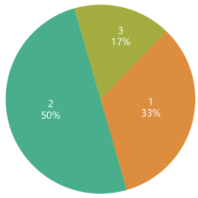
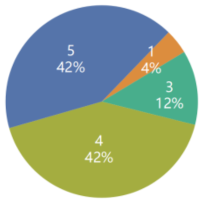
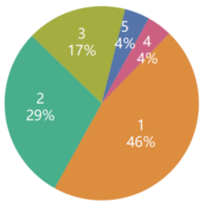
5	I found the various functions in the system were well integrated.	(score-1)*2.5=8.65	
6	I thought there was too much inconsistency in this system.	(5-score)*2.5=8.23	
7	I would imagine that most people would learn to use this system very quickly.	(score-1)*2.5=8.13	
8	I found the system very cumbersome to use.	(5-score)*2.5=7.92	
9	I felt very confident using this system.	(score-1)*2.5=7.92	
10	I needed to learn a lot of things before I could get going with this system.	(5-score)*2.5=7.71	
SUM UP		80.33	

Table 3: Usability Testing

According to the data above, we can indicate that people like our product and will recommend it to their friends for its score is above 80.3, which is considered an A. Meanwhile, the form that has been shown above also exposes many problems with our prototype. Firstly, we found that the score of the second question was meagre compared with the others. It means that users thought our project has unnecessarily complex. To work out this problem, the main subject of the prototype was defined again, and some non-essential features were removed based on the discussion results. Besides, scores for the third and last questions were relatively low too. Both showed us that users could not have a good command of how to use it at the beginning. Therefore, we set up a "help" option to help them get familiar with the project quicker.

Surprisingly, many respondents gave us valuable suggestions after finishing the questionnaire. For example, one of the participants suggested we readjust the position of the title on the "Sports" page to present a better visual effect. We adopt this advice and found the balance of the pages has been dramatically improved.

All in all, we think this evaluation is meaningful for the prototype. In the beginning, we thought our design did not have any problems. However, when it came to the actual test phrase, we found many problems we had not expected. Thanks to the evaluation, we can receive much objective advice, which is critical for our product.

5 Conclusion

5.1 Summary

In summary, this project has introduced a comprehensive music assessment and recommendation system tailored to the needs of adults pursuing musical instrument learning. The system emphasizes multi-dimensional evaluation, precision in comparisons, personalized recommendations, and user feedback-driven optimization. It addresses the challenges posed by the music education industry, offering a user-centric approach for improving the learning experience.

5.2 Limitation

Despite its strengths, the system has certain limitations. One notable limitation is the reliance on textual input for music recommendations, which may not fully capture users' preferences. Additionally, the accuracy of music similarity calculations may be affected by the diversity of music genres and cultural aspects. Noise reduction in audio analysis remains a challenge, especially in noisy environments. Furthermore, the system's effectiveness may vary based on the availability and quality of training data.

5.3 Future Plan

In the future, the system can undergo enhancements by incorporating advanced natural language processing techniques to gain a deeper understanding of user preferences. Efforts may also be directed toward improving the accuracy of music similarity calculations and addressing the challenges posed by diverse music genres and cultural factors. Further refinement of noise reduction techniques will be undertaken to ensure accurate analysis in various acoustic environments.

However, we understand that the journey of learning a musical instrument is not static.

It's dynamic, and it evolves with the learner's progress. Thus, Therefore, optimization driven by user feedback is a part of our system that needs to be improved and supplemented. We value the input of our users and actively seek their feedback to continually refine and enhance the system's functionality, ensuring that it evolves in harmony with the needs and expectations of our users.

Additional experiments will be conducted to evaluate the performance of OpenL3 in various audio similarity tasks. Moreover, the exploration of alternative feature extraction methods and similarity metrics will be a focus, aiming to enhance the precision of audio similarity computations.

Expanding the training data and refining the underlying model will be pivotal in augmenting the system's overall performance and reliability. Lastly, the integration of real-time user interaction and comprehensive user feedback mechanisms will serve as a catalyst for continuous system improvements.

6 References

[1] Nakamura, E., Yoshii, K. and Katayose, H., 2017. Performance Error Detection and Post-Processing for Fast and Accurate Symbolic Music Alignment. In ISMIR (pp. 347-353).
[2] High-resolution Piano Transcription with Pedals by Regressing Onsets and Offsets Times, [To appear], 2020
[3] Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J., & Serra, X. (2013). Essentia. Proceedings of the 21st ACM International Conference on Multimedia. https://doi.org/10.1145/2502081.2502229
[4] Serra, J., Serra, X., & Andrzejak, R. G. (2009). Cross recurrence quantification for cover song identification.New Journal of Physics.
[5] Serra, Joan, et al (2008). Chroma binary similarity and local alignment applied to cover song identification. IEEE Transactions on Audio, Speech, and Language Processing.
[6] Serra, J., Gómez, E., & Herrera, P. (2008). Transposing chroma representations to a common key, IEEE Conference on The Use of Symbols to Represent Music and Multimedia Objects.
[7] R. Arandjelovic and A. Zisserman, “Look, listen and learn,” in Proceedings of the IEEE International Conference on Computer Vision, 2017.
[8] J. Cramer, H.-H. Wu, J. Salamon, and J. P. Bello, “Look, listen, and learn more: Design choices for deep audio embeddings,” in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 3852–3856.
[9] Kong, Q., Li, B., Chen, J., & Wang, Y. (2022, May 12). Giantmidi-piano: A large-scale MIDI dataset for classical piano music. Transactions of the International Society for Music Information Retrieval.
[10] Cramer, J., Jordà, S., & Peeters, G. (2017). A large set of audio features for sound description (similarity and classification) in the CUIDADO project. Zenodo. https://zenodo.org/record/2552868
[11] Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Bittner, R., Wu, S., ... & Eck, D. (2019). Enabling factorized piano music modeling and generation with the MAESTRO dataset. arXiv preprint arXiv:1912.01637.

7 Appendix

7.1 Appendix of report: Project Proposal

GRADUATE CERTIFICATE: Intelligent Reasoning Systems (IRS)
PRACTICE MODULE: Project Proposal

Date of proposal: 26 September 2023
Project Title: Unlocking the Gateway to Music: A New Assistant for Music Learning
Sponsor/Client: *(Name, Address, Telephone No. and Contact Name)* Institute of Systems Science (ISS) at 25 Heng Mui Keng Terrace, Singapore NATIONAL UNIVERSITY OF SINGAPORE (NUS) Contact: Mr. GU ZHAN / Lecturer & Consultant Telephone No.: 65-6516 8021 Email: zhan.gu@nus.edu.sg
Background/Aims/Objectives: The proposed intelligent eco-system will make use of various advanced machine reasoning techniques and components to help people to learn guitar.
Requirements Overview: • Research ability • Programming ability • System integration ability
Resource Requirements (please list Hardware, Software and any other resources) Hardware proposed for consideration: • GPU, etc. Software proposed for consideration: • Reasoning systems, e.g. SVM, Collaborative Filtering, Optimization, etc • Pertained machine learning models, e.g. Vision, Speech, NLP

- Machine learning use cases, e.g. Orange3, R
- Deep learning tools, e.g. Neural Network Console Sony, Python Keras
- Chat-bots, e.g. ChatterBot, DBpedia Chat-bot
- Cognitive systems, e.g. MyCroft
- Robotic Process Automation, .e.g TagUI
- Cloud computing/server, e.g. Amazon, Google, IBM, Azure, etc.
- Application container, e.g. Docker

Number of Learner Interns required: (Please specify their tasks if possible)

a team of three project members

Methods and Standards:

Procedures	Objective	Key Activities
Requirement Gathering and Analysis	The team should meet with ISS to scope the details of project and ensure the achievement of business objectives.	1. Gather & Analyze Requirements 2. Define internal and External Design 3. Prioritize & Consolidate Requirements 4. Establish Functional Baseline
Technical Construction	· To develop the source code in accordance to the design. · To perform unit testing to ensure the quality before the components are integrated as a whole project	1. Setup Development Environment 2. Understand the System Context, Design 3. Perform Coding 4. Conduct Unit Testing
Integration Testing and acceptance testing	To ensure interface compatibility and confirm that the integrated system hardware and system software meets requirements and is ready for acceptance testing.	1. Prepare System Test Specifications 2. Prepare for Test Execution 3. Conduct System Integration Testing 4. Evaluate Testing 5. Establish Product Baseline
Acceptance Testing	To obtain ISS user acceptance that the system meets the requirements.	1. Plan for Acceptance Testing 2. Conduct Training for Acceptance Testing 3. Prepare for Acceptance Test Execution 4. ISS Evaluate Testing 5. Obtain Customer Acceptance Sign-off

Delivery	To deploy the system into production (ISS standalone server) environment.	1. Software must be packed by following ISS's standard 2. Deployment guideline must be provided in ISS production (ISS standalone server) format 3. Production (ISS standalone server) support and troubleshooting process must be defined.

Team Formation & Registration

Team Name: GROUP16
Project Title (repeated): Unlocking the Gateway to Music: A New Assistant for Music Learning
System Name (if decided):
Team Member 1 Name: Liu Jingyi
Team Member 1 Matriculation Number: A0285813X
Team Member 1 Contact (Mobile/Email): e1221625@u.nus.edu
Team Member 2 Name: Yi Ying

Team Member 2 Matriculation Number: A0285683M
Team Member 2 Contact (Mobile/Email): ying.yi@u.nus.edu
Team Member 3 Name: Zhang Ni
Team Member 3 Matriculation Number: A0285674M
Team Member 3 Contact (Mobile/Email): zhangniandhe@outlook.com

For ISS Use Only		
Programme Name:	**Project No:**	**Learner Batch:**
Accepted/Rejected/KIV:		
Learners Assigned:		

Advisor Assigned:

Contact: Mr. GU ZHAN / Lecturer & Consultant
Telephone No.: 65-6516 8021
Email: zhan.gu@nus.edu.sg

7.2 Appendix of report: Mapped System Functionalities against knowledge, techniques, and skills of modular courses: MR, RS, CGS

In the proposed project, we have integrated and demonstrated multiple aspects from the following four technique groups:

1. Decision Automation: Business Rules & Process

We perform traditional similarity calculation through collaborative filtering, utilizing business rules to assess the similarities among music compositions.
For piano recognition, we employ deep learning methods using Convolutional Neural Networks (CNN), which are governed by knowledge-based reasoning techniques, particularly in assessing and recognizing piano music.

2. Business Resource Optimization: Informed Search

We use crawler technology to acquire MP3 files from YouTube, thereby optimizing our resources for building a comprehensive testing database. This informed search aids in data collection and resource management.

3. Knowledge Discovery & (Big) Data Mining Techniques

In the recommendation system, we incorporate knowledge discovery techniques by employing the collaborative filtering algorithm, specifically item-k-nearest neighbors (item-kNN). This method measures the similarity between user input and the music database, providing recommendations based on knowledge mining.
We further utilize the TF-IDF method to preprocess music information by converting textual data into vector representations. This process leverages data mining techniques to assess term importance within the entire corpus.

4. System Designed with Cognitive Techniques or Tools

Our system integrates cognitive techniques, including knowledge graph and natural language processing (NLP). These components enhance the system's capabilities for human communication. We employ NLP for generating user-friendly text comments

and suggestions, enabling effective user interaction.
We also incorporate personalized suggestions based on individual user data, demonstrating cognitive techniques that facilitate tailored content delivery in alignment with user performance and learning needs.

Overall, our system is intricately designed to encompass elements from these four technique groups, ensuring that it meets market demands and user expectations. Through the utilization of various techniques, such as audio analysis and natural language processing, we offer high-quality, personalized recommendations to our users, all while maintaining flexibility and adaptability through continuous improvement driven by user feedback.

7.3 Appendix of report: Installation and User Guide

7.3.1 Installation

Python 3.10.9
Essentia
numpy 1.16.3
librosa==0.7.2
Piano_ Translation_ Inference
Beautiful Soup 4.0
matplotlib 3.1.1
torch 1.9.0
libsndfile1
beautifulsoup4==4.9.1
nltk==3.5
piano_transcription_inference
youtube-dl
PyTorch (>=1.4)

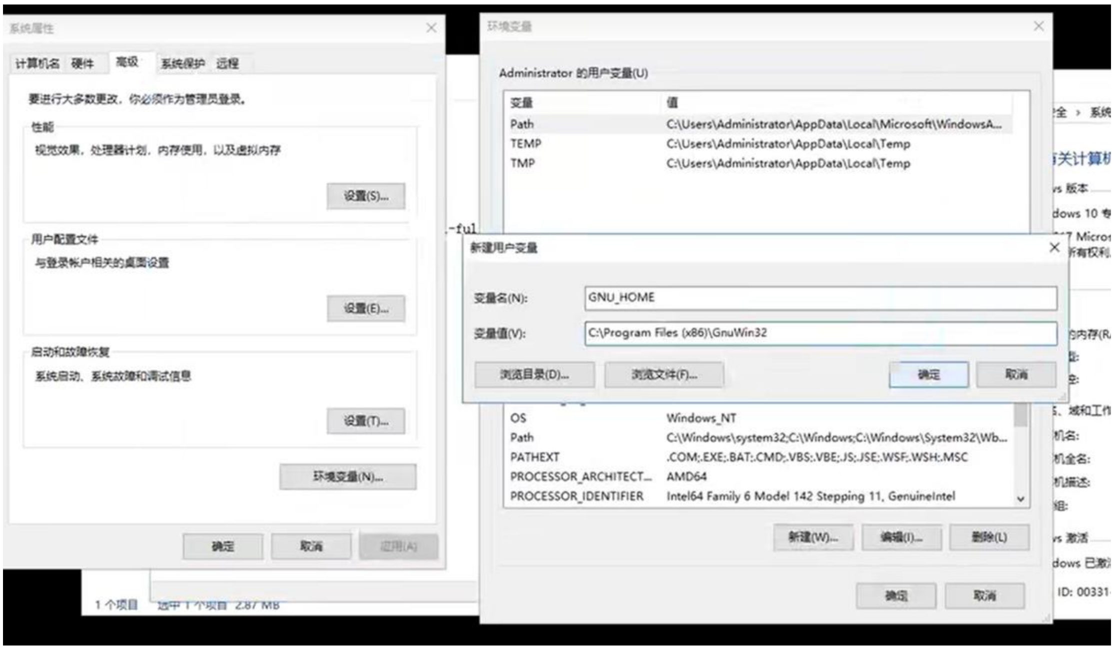
7.3.2 User Guide

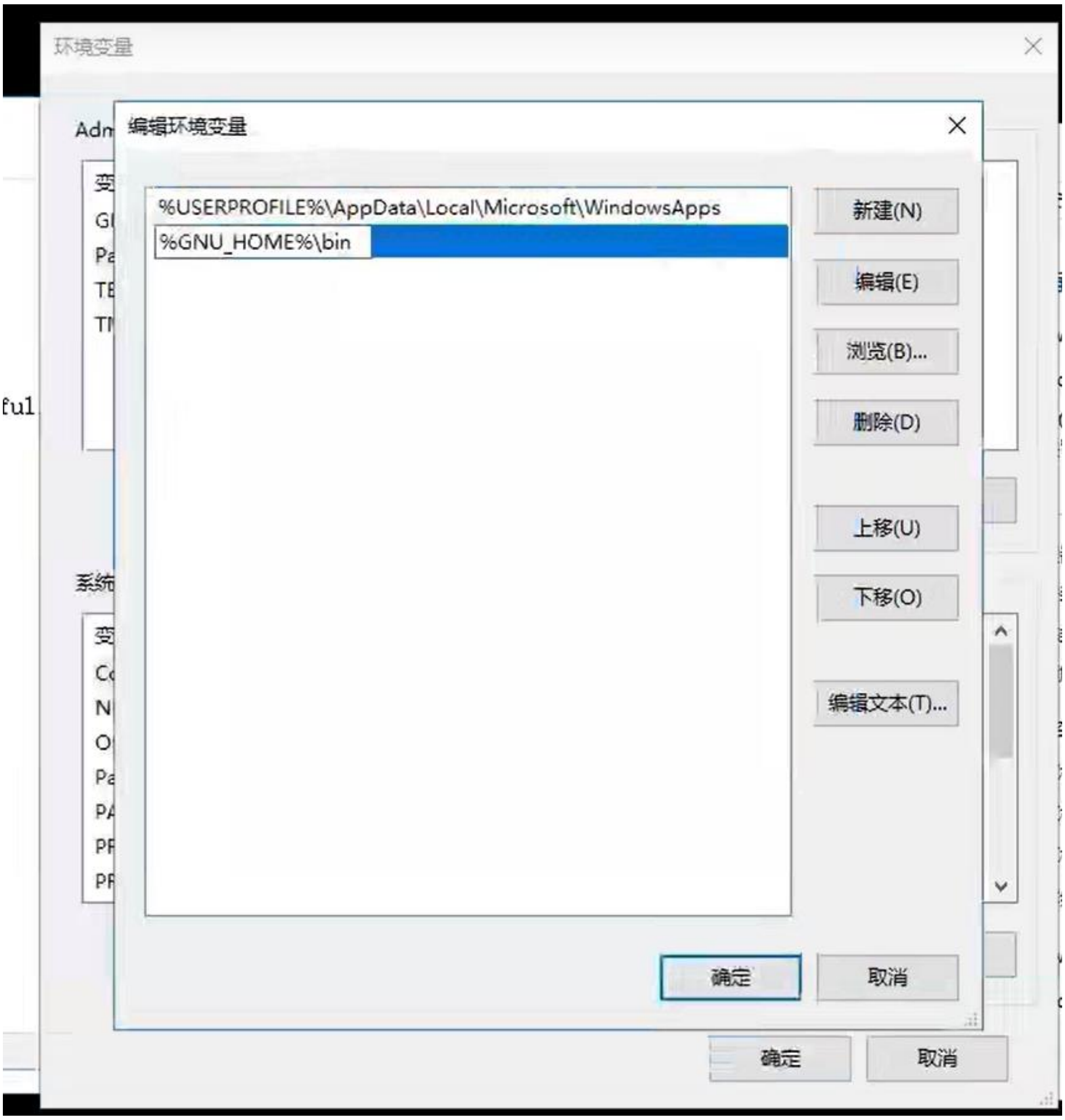
Install PyTorch (>=1.4) following https://pytorch.org/.

pip install -r requirements.txt

Install the wget.

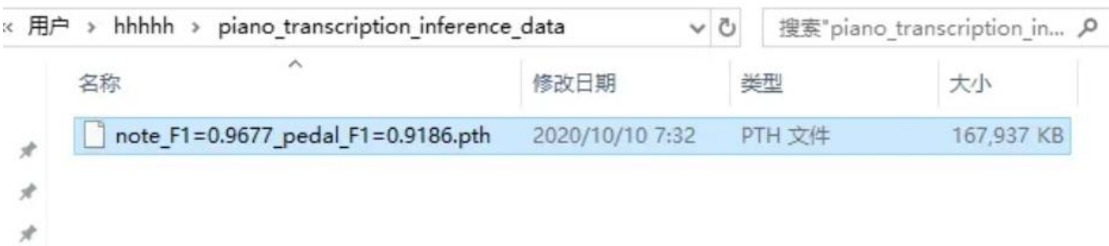
Add system path.



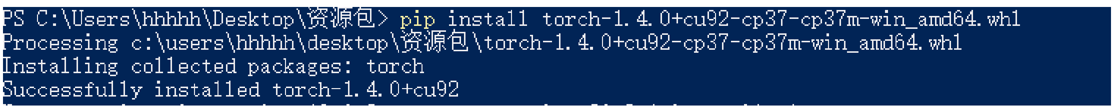


Place the file note_F1=0.9677_pedal_F1=0.9186.pth in the following directory: (Create if not found)

C:\Users\your account name\piano_transcription_inference_data\



Install python and pytorch.



Install ffmpeg

Move the folder ffmpeg-4.3.1-2020-10-01-full_build to the directory
C:\Program Files\ Below

Install requirements.txt:
pip install -r requirements.txt

```
    100% |████████████████████████████████| 327kB 1.3MB/s
Requirement already satisfied: setuptools in c:\users\administrator\appdata\local\programs\python\python37\lib\site-packages (from numba==0.48->-r requirements.txt (line 4)) (40.8.0)
Collecting llvmlite<0.32.0,>=0.31.0dev0 (from numba==0.48->-r requirements.txt (line 4))
  Downloading https://pypi.doubanio.com/packages/1d/83/cd2843726a6316e372822e9e42cd0083b6d1d98d89d53880e7e67d5eec68/llvmlite-0.31.0-cp37-cp37m-win_amd64.whl (13.6MB)
    100% |████████████████████████████████| 13.6MB 1.2MB/s
Collecting future (from mir_eval==0.5->-r requirements.txt (line 6))
  Downloading https://pypi.doubanio.com/packages/45/0b/38b06fd9b92dc2b68d58b75f900e97884c45bedd2ff83203d933cf5851c9/future-0.18.2.tar.gz (829kB)
    100% |████████████████████████████████| 829kB 984kB/s
Collecting cycler>=0.10 (from matplotlib==3.0.3->-r requirements.txt (line 7))
  Downloading https://pypi.doubanio.com/packages/f7/d2/e07d3ebb2bd7af696440ce7e754c59dd546ffe1bbe732c8ab68b9c834e61/cycler-0.10.0-py2.py3-none-any.whl
Collecting kiwisolver>=1.0.1 (from matplotlib==3.0.3->-r requirements.txt (line 7))
  Downloading https://pypi.doubanio.com/packages/b2/55/6681ac2cc8de9bb612b1a777606e5beef240bf63aaa6cb03f44af5f42a77/kiwisolver-1.3.1-cp37-cp37m-win_amd64.whl (51kB)
    100% |████████████████████████████████| 61kB 8.0MB/s
Collecting pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 (from matplotlib==3.0.3->-r requirements.txt (line 7))
  Downloading https://pypi.doubanio.com/packages/8a/bb/488841f56197b13700afd5658fc279a2025a39e22449b7cf29864669b15d/pyparsing-2.4.7-py2.py3-none-any.whl (67kB)
    100% |████████████████████████████████| 71kB 408kB/s
Collecting threadpoolctl>=2.0.0 (from scikit-learn!=0.19.0,>=0.14.0->librosa==0.6.0->-r requirements.txt (line 3))
  Downloading https://pypi.doubanio.com/packages/f7/12/ec3f2e203afa394a149911729357aa48affc59c20e2c1c8297a60f33f133/threadpoolctl-2.1.0-py3-none-any.whl
Installing collected packages: numpy, six, h5py, python-dateutil, pytz, pandas, audioread, scipy, joblib, threadpoolctl,
 scikit-learn, decorator, llvmlite, numba, resampy, librosa, mido, future, mir-eval, cycler, kiwisolver, pyparsing, matplotlib, torchlibrosa, sox
  Running setup.py install for audioread ... done
```

After the installation is complete, restart to ensure that all settings take effect.

Use pip to install piano_transcription_inference and run it.

```
In [1]: from piano_transcription_inference import PianoTranscription, sample_rate, load_audio

In [2]: # Load audio
        (audio, _) = load_audio('./Haru_Haru.mp3', sr=sample_rate, mono=True)

In [6]: # Transcriptor
        transcriptor = PianoTranscription(device='cpu', checkpoint_path=None)     # 'cuda' | 'cpu'

        Checkpoint path: C:\Users\60963/piano_transcription_inference_data/note_F1=0.9677_pedal_F1=0.9186.pth
        Using cpu for inference.
        Using CPU.

In [4]: # Transcribe and write out to MIDI file
        transcribed_dict = transcriptor.transcribe(audio, 'Haru_Haru.mid')

        Segment 0 / 17
        Segment 1 / 17
        Segment 2 / 17
        Segment 3 / 17
        Segment 4 / 17
        Segment 5 / 17
        Segment 6 / 17
        Segment 7 / 17
        Segment 8 / 17
        Segment 9 / 17
        Segment 10 / 17
        Segment 11 / 17
        Segment 12 / 17
        Segment 13 / 17
        Segment 14 / 17
        Segment 15 / 17
        Segment 16 / 17
        Segment 17 / 17
        Write out to Haru_Haru.mid
```

7.4 Appendix of report: 1 or 2 pages individual project report per project member

7.4.1 Individual Report – Yi Ying

Name: Yi Ying	**Matriculation Number:** A0285683M

Personal contribution to group project:

My contributions to the group project encompassed a wide range of responsibilities. They are as follows:

1. Project Ideation and Conceptualization
   I actively participated in brainstorming sessions and proposed the project's initial idea. My input was instrumental in defining the scope and objectives of the project.
2. Team Leader
   I assumed a leadership role in guiding the technical aspects of the project. This involved designing the system architecture, which served as the backbone of our software solution.
3. System Architecture and Overall Design
   I provided a foundational framework that allowed the project to successfully incorporate various technical elements and deliver a comprehensive solution.
4. User Interface (UI) Design
   A key aspect of our project was creating a visually appealing and user-friendly interface. I designed the Use Case Demo, ensuring that it was both aesthetically pleasing and functional.
5. Data Management and Processing
   I took charge of organizing the data collection process and managing the project's database. Additionally, I oversaw the conversion of MP3 files into MIDI files and the extraction of essential audio features.
6. AI Implementation
   Leveraging my understanding of artificial intelligence, I Implemented audio feature extraction, using OpenL3 preprocessing model, which is crucial to improving the project's music evaluation capabilities, and completed multi-dimensional scoring.
7. Similarity Calculation
   I was responsible for developing the similarity calculation module, which required intricate mathematical operations and analysis.
8. Documentation
   I played a crucial role in writing the project report and amalgamated the project report, ensuring that it effectively conveyed our project's goals, methodologies, and outcomes. Completed the report outline, system architecture and overall design, model development, system evaluation, summary sections and the structure diagram of the project.
9. Project Integration
   My responsibilities included harmonizing team efforts, ensuring that the project's different components aligned cohesively.
10. Project Planning and Coordination
    I assumed a central role in planning the project and coordinating tasks and resources among team members.

11. Project Testing and Evaluation

Ensuring that the project met our envisioned standards was a key part of my responsibilities.

What learnt is most useful for you:

Through my involvement in this project, I have gained invaluable insights and practical experience in various domains. Firstly, the process of project ideation and conceptualization has underscored the significance of clearly defining project scope and objectives, which has significantly fortified my technical foundation. Managing a project involves a multitude of tasks, including the quest for suitable datasets, configuring, and setting up the computing environment, and resolving potential conflicts between various tools. Throughout this journey, I familiarized myself with several Python libraries, such as librosa, numpy, and pandas, among others.

The process of building a model has been a transformative learning experience. It has imparted valuable lessons on how to explore, experiment, and innovate within the realm of data and algorithms. This experience has underscored the importance of delving deep into problem understanding and selecting the most suitable model that aligns with the data and project objectives. It has also emphasized the critical role of data preprocessing, feature engineering, and model evaluation in the pursuit of optimal performance.

Furthermore, my engagement in this project has honed my technical leadership skills, particularly in the domains of system architecture and overall design, which are essential for constructing cohesive and efficient software solutions. Selecting the most appropriate model for solving practical application problems is a challenge that requires careful consideration.

Additionally, my exposure to user interface (UI) design has provided me with the ability to strike a balance between aesthetics and functionality. The responsibilities I assumed in data management and processing have strengthened my capabilities in data organization and analysis.

In terms of artificial intelligence (AI), my experience with AI implementation and similarity calculations has broadened my knowledge of AI and its diverse applications. Finally, my roles in project planning, coordination, and testing have equipped me with comprehensive project management skills.

How you can apply the knowledge and skills in other situations or your workplaces:

The knowledge and skills I have gained through this project are highly versatile and applicable in diverse professional settings. The expertise in project ideation and conceptualization is valuable in any scenario where the clear definition of goals and objectives is pivotal. My proficiency in technical leadership, system architecture, and design, along with UI design, can prove advantageous in software development projects spanning various industries. For roles centered around data, such as data analysis and database management, my data management and processing skills are invaluable. Additionally, my knowledge of AI implementation and similarity calculations finds relevance in the expansive domains of artificial intelligence and machine learning. Lastly, the project management and coordination competencies I have acquired can be leveraged in any team-based project or managerial role. In summary, the diverse skill set I have cultivated throughout this project equips me to excel in numerous work environments, positioning me as a valuable asset in various professional contexts.

7.4.2 Individual Report – Jingyi Liu

Name: Jingyi Liu	**Matriculation Number:** A0285813X

Personal contribution to group project:

1. Ideation of the project's structure
2. Piano music database creation
   I took charge of creating and organizing the piano music database. By crawling different kinds of piano music information from RenRen Piano Music net, I ensured the system had a structured and sufficient dataset for accurate recommendations.
3. Business value research
   By analyzing market situation, business profit, trends and defect of other Competitors, I identified the potential and business value for generation of this piano music learning assistant system.
4. Creation of whole recommendation system
   Developed a system algorithm capable of recommending piano music from the dataset based on user preferences. By building the Item-KNN model from the given music database, we can use it to predict a recommendation list that is most similar to the user input.
5. Implemented a tf-idf preprocessing method for the music data
   I wrote the function to transform the music data into the vector that machine learning phase needs.
6. Reporting and Documentation
   I also helped with writing comprehensive documentation for this project. This included Background, Project Objectives, Market Research, Problem Description and Competitors Analysis etc.

What learnt is most useful for you:

1. learning how to get data

One of the challenges encountered in the project was the lack of piano music database. By tacking this issue, I learned how to crawl data from the web by myself and applied the skill to crawling piano music data.

2. learning how to deal with chinese characters in preprocessing

Another challenge is that TfidfVectorizer can not convert the chinese characters to the vectors properly. Therefore, I had a chance to acquire some tricks to tackle this problem by searching relevant references.

3. Applying and Understanding item-knn Algorithms

Through my generation of the recommendation system, I delved deeper into similarity algorithms such as cosine similarity, K-nearest neighbors (KNN). Building the item-knn model successfully in this project successfully allowed me to gain hands-on experience and deeper understanding in the machine-learning skills that we learned in class.

How you can apply the knowledge and skills in other situations or your workplaces:

As for getting data , I gained the skills and experience to crawl data from a website. this can be applied in many projects because a lot of current artificial intelligent work requires crawling data to build a broaden data sets.

And I also understand how to update setting of TfidfVectorizer to fit the chinese characters well. That skill is quite useful in NLP preprocessing area.

Finally, The item based recommendation method performed in this project can also be applied to solve many real-world problems. In fact, the project reveals a method to recommend items that fit people's preferences, and it could help people discover things they like and create more profits for companies.

7.4.3 Individual Report – Zhang Ni

Name: Zhang Ni	**Matriculation Number:** A0285674M

Personal contribution to group project:

1. Acquisition of database test data:

My research first involves obtaining test data from YouTube for music similarity analysis. To achieve this goal, I used the crawler technology in the Python programming language. Specifically, I used third-party libraries such as Beautiful Soup and Requests to access and download MP3 audio files on YouTube. These libraries have helped me automate the process of data collection, greatly improving efficiency.

Once I have obtained MP3 audio files, I also need to convert them into MIDI files for subsequent audio analysis.

2. CNN model training for piano recognition:

Piano recognition is the core part of my research. I trained a Convolutional Neural Network (CNN) model that can recognize piano notes in audio. The architecture of the model includes convolutional layer, pooling layer, and fully connected layer. I used a large number of labeled datasets, including note position and pitch information, to train this model.

During the training process, I used optimization algorithms such as random gradient descent and adjusted different hyperparameters, such as learning rate and batch size. To evaluate the performance of the model, I used standard classification indicators such as accuracy and recall. These indicators helped me determine the accuracy and reliability of the model.

3. Collaborative filtering similarity comparison:

In addition to deep learning models, I also use collaborative filtering algorithms for music similarity comparison. This method relies on the interaction data between users and music, and recommends music to users by calculating the similarity between users. I implemented user based collaborative filtering and used an open source recommendation system library to simplify this process.

4. Audio visualization through Essentia:

In order to present the data more vividly, I used Essentia, an audio processing tool. Through Essentia, I am able to convert user input audio into visual images, including audio spectrum, pitch

contours, and more. These images provide more information and help users better understand music features.
Audio visualization also helps me with data analysis. This helps me to delve deeper into the characteristics of music in my research.
5. Personalize recommendation creation
Use the OPENAI API to create the comment for users, including the suggestions that can help user to improve their skills and the songs that are suitable for users to practice.
6. PPT design and video making :
I have designed the proposal presentation and the two videos to effectively convey our research goals and methods in both academic and professional fields. This PPT includes the structure of the slide, such as title, background, goal, etc. I choose to use charts, images, and logos to enhance the visual appeal of the slides. In addition, I ensure that the content outline of the PPT and videos are clear and logically coherent.
7. Filling out relevant reports:
When preparing the relevant report, I first constructed the structure of the report, including the introduction, methods, results, and discussion sections.
8.Writing proposal
Integrate project information and give the detail of our project.
9.Inprove the project structure
Searching the existing tools and give the idea about how to get the score of the music, like how to get test dataset, making piano detection, and using collaborative filtering algorithms for music similarity comparison.