International Journal on Science and Technology (IJSAT)

E-ISSN: 2229-7677 ● Website: www.ijsat.org ● Email: editor@ijsat.org

An Ai-Driven Approach for Comprehensive Song Feedback: From Audio Processing to Quality Enhancement

Aarthi Mai A S[1], Rajeswari N[2]

[1]Assistant Professor, Department of Computer Applications, PSG College of Technology, Coimbatore
[2]Assistant Professor (Senior Grade), Department of Computer Applications, PSG College of Technology, Coimbatore
[1]asa.mca@psgtech.ac.in, [2]nrj.mca@psgtech.ac.in

Abstract:

This paper presents a comprehensive approach to developing an AI algorithm designed to provide detailed feedback on input songs. The methodology comprises several key stages: preprocessing, feature extraction, analysis, feedback generation, and output. Initially, the preprocessing phase involves loading the audio, converting it to mono, and resampling to ensure uniformity. Feature extraction follows, capturing time-domain features such as tempo and beats, frequency-domain features including spectral centroid and Mel-frequency cepstral coefficients (MFCCs), rhythm patterns, and harmonic content via chroma features. Subsequent analysis utilizes emotion recognition, genre classification, and quality assessment techniques to evaluate the song's characteristics. The feedback generation phase involves comparing the extracted features with reference data, identifying the song's strengths, and suggesting improvements. Finally, the feedback is formatted and recommendations are provided in a structured manner to assist in enhancing the song's quality.

KEYWORDS: AI Algorithm, Audio Processing, Feature Extraction, Emotion Recognition, Genre Classification, Quality Assessment, Feedback Generation

1. INTRODUCTION

The advent of artificial intelligence has transformed numerous fields, including music analysis and feedback. This paper introduces a novel AI algorithm tailored to deliver in-depth feedback on input songs through a structured and multi-phase approach. The methodology outlined encompasses several critical stages:

Preprocessing, Feature Extraction, Analysis, Feedback Generation, and Output.

The preprocessing phase establishes a uniform foundation by loading the audio files, converting them to mono, and resampling them to a consistent rate. This ensures that subsequent analyses are conducted on standardized data. During the feature extraction stage, the algorithm gathers a wide array of audio features, including time-domain elements like tempo and beats, frequency-domain characteristics such as spectral centroid and Mel-frequency cepstral coefficients (MFCCs), along with rhythm patterns and harmonic content through chroma features.

The analysis phase leverages advanced techniques in emotion recognition, genre classification, and quality assessment to comprehensively evaluate the song's attributes. This analysis informs the feedback generation phase, where the algorithm compares the extracted features against reference data to identify the song's strengths and areas for improvement. The final stage focuses on presenting the feedback and recommendations in a clear, actionable format to help improve the overall quality of the song.

By integrating these stages, this approach provides a robust framework for generating insightful and constructive feedback, thereby contributing to the advancement of music analysis and AI-assisted music enhancement.

2. ALGORITHM FOR COMPREHENSIVE SONG FEEDBACK

The steps involved in developing AI algorithm to give feedback on a given input song involves multiple components, including audio processing, feature extraction, and machine learning techniques.

STEP 1: PREPROCESSING THE GIVEN SONG

Preprocessing audio data is a crucial step in many audio processing and machine learning applications. By standardizing the input data, we can ensure consistency and reliability in subsequent analysis or modeling tasks. The fundamental steps of preprocessing an audio file, focusing on loading the song, converting it to mono, and resampling the audio to a consistent sample rate using the librosa library are as follows.

a. **Load the Song**: Loading an audio file involves reading the file into a format that can be manipulated and analyzed programmatically. The librosa library in Python is a powerful tool for this purpose, as it provides simple functions to load audio files in various formats.

Implementation: To load a song using librosa, typically use the librosa.load function, which reads the audio file and returns the audio time series and the sample rate.

```
import librosa
file_path = 'path/to/your/audio/file.mp3'
audio_data, sample_rate = librosa.load(file_path)
```

b. Convert to Mono: Stereo audio contains two channels (left and right), which can add complexity to the analysis. For many applications, such as machine learning models, it is beneficial to convert stereo audio to mono (single channel) to simplify processing and reduce computational load.

Implementation: The librosa.load function has an argument mono which, when set to True, converts the audio to mono automatically.

```
audio_data, sample_rate = librosa.load(file_path, mono=True)
```

c. Resample: Audio files can have different sample rates, which is the number of samples of audio carried per second. Consistent sample rates are essential for uniform processing, ensuring that all audio data is treated similarly in subsequent steps.

Implementation: The librosa.resample function can be used to change the sample rate of the audio data to a desired target sample rate.

```
target_sample_rate=22050 audio_data=librosa.resample(audio_data,
orig_sr=sample_rate, target_sr=target_sample_rate)
sample_rate = target_sample_rate
```

By following these preprocessing steps—loading the song, converting to mono, and resampling—prepare the audio data for further analysis or processing tasks. These steps ensure that the input data is in a consistent and manageable format, paving the way for reliable and efficient audio processing workflows.

STEP 2: FEATURE EXTRACTION

Feature extraction is a crucial step in audio processing and music information retrieval. It involves converting raw audio data into a set of features that can be used for various tasks such as classification, recognition, and analysis. These features can be categorized into different domains based on the type of information they represent. In this step, let us explore time-domain features, frequency-domain features, rhythm features, and harmony features.

a. Time-Domain Features: Time-domain features are extracted directly from the audio signal in the time domain. These features represent how the audio signal varies over time.

- **Tempo**: Tempo is the speed at which a piece of music is played, usually measured in beats per minute (BPM). It is a critical feature for identifying the genre, mood, and style of a musical piece.
- **Beats**: The beat is the basic unit of time in a piece of music, representing the regular rhythmic pulse. Detecting beats helps in understanding the rhythmic structure of the music.
- **Zero-Crossing Rate (ZCR)**: The zero-crossing rate is the rate at which the audio signal changes sign from positive to negative or vice versa. It is a simple feature that can indicate the noisiness or the presence of percussive sounds in the audio.

b. Frequency-Domain Features : Frequency-domain features are extracted by transforming the audio signal from the time domain to the frequency domain using techniques like the Fourier Transform. These features capture the spectral characteristics of the audio.

- **Spectral Centroid**: The spectral centroid indicates the center of mass of the spectrum and is often associated with the perceived brightness of a sound. It is calculated as the weighted mean of the frequencies present in the signal.
- **Spectral Bandwidth**: The spectral bandwidth measures the width of the spectrum, indicating the range of frequencies present in the signal. It can provide information about the timbral characteristics of the audio.
- **Mel-Frequency Cepstral Coefficients (MFCCs)**: MFCCs are a representation of the short-term power spectrum of a sound. They are widely used in speech and audio processing because they approximate the human ear's response more closely than other features. MFCCs are derived from the logarithm of the mel-scaled power spectrum of the audio signal.

c. Rhythm Features : Rhythm features capture the temporal patterns and rhythmic structure of the music.

- **Rhythm Patterns**: Rhythm patterns are repetitive sequences of beats or other rhythmic elements in a piece of music. Extracting rhythm patterns helps in identifying the style and structure of the music.
- **Beat-Related Features**: Beat-related features include information about the timing, strength, and variation of beats in the music. These features are crucial for tasks like beat tracking and tempo estimation.

4. Harmony Features: Harmony features capture the harmonic content and the relationship between different pitches in the music.

- **Chroma Features**: Chroma features represent the 12 different pitch classes (semitones) of the musical octave, disregarding the octave in which the pitch appears. These features are useful for analyzing harmonic content, chord recognition, and key detection. They provide a compact representation of the harmonic structure of the music.

Feature extraction in audio processing is essential for converting raw audio data into meaningful representations that can be used for various analytical and machine learning tasks. By extracting time-domain, frequency-domain, rhythm, and harmony features, we can gain a comprehensive understanding of the audio signal's characteristics and use this information for further processing and analysis.

STEP 3: ANALYZE THE FEATURES

Analyzing the features of a song involves delving into various aspects that define its characteristics and overall impact. Here are three key areas of feature analysis: Emotion Recognition, Genre Classification, and Quality Assessment.

a. Emotion Recognition : Emotion recognition in music aims to identify the emotional content conveyed by a song. This involves understanding how different musical elements evoke specific emotions in listeners.

- **Pre-trained Models:** These are machine learning models that have already been trained on large datasets to recognize patterns. For emotion recognition, these models are typically trained on datasets labeled with various emotions.
- **Emotional Content:** The feelings or emotional response a piece of music evokes, such as happiness, sadness, excitement, or calmness.
- **Musical Elements:** Components like melody, harmony, rhythm, tempo, and dynamics contribute to the emotional tone of a song.

How It Works:

1. **Input Data:** The song's audio features, such as pitch, rhythm, and tempo, are extracted.
2. **Model Application:** A pre-trained emotion recognition model analyzes these features to predict the dominant emotion(s) expressed in the song.
3. **Output:** The model outputs labels or scores corresponding to different emotions (e.g., happy, sad, energetic, calm).

b. Genre Classification: Genre classification involves categorizing music into different genres based on its characteristics and stylistic elements.

- **Genre:** A category that defines the style of a song, such as rock, jazz, classical, pop, or hip-hop.
- **Classification Models:** Machine learning models specifically trained to recognize and differentiate between various music genres.
- **Feature Extraction:** The process of extracting relevant audio features (e.g., timbre, rhythm, tempo) that help in identifying the genre.

How It Works:

1. **Input Data:** The song's audio features are extracted and pre-processed.
2. **Model Application:** A genre classification model analyzes these features and assigns the song to one or more genres.
3. **Output:** The model provides a genre label (or multiple labels in case of multi-genre songs).

3. Quality Assessment : Quality assessment in music focuses on evaluating the technical and aesthetic quality of a song. This involves analyzing aspects like clarity, balance, and loudness.

- **Clarity:** The clearness and distinctness of the audio, free from noise and distortion.
- **Balance:** The even distribution of audio elements across different frequencies and channels.
- **Loudness:** The perceived volume of the song, which should be consistent and within acceptable levels to avoid listener fatigue.

How It Works:

1. **Input Data:** The audio signal is analyzed to extract features related to clarity, balance, and loudness.
2. **Model Application:** Quality assessment models evaluate these features against standard metrics.
3. **Output:** The model provides scores or ratings for each quality aspect, indicating the overall technical quality of the song.

Analyzing the features of a song through emotion recognition, genre classification, and quality assessment provides a comprehensive understanding of its characteristics and impact. By leveraging pre-trained models and advanced audio analysis techniques, we can gain insights into the emotional tone, stylistic genre, and technical quality of the music. This process is crucial for various applications in music production, recommendation systems, and musicology research.

STEP 4: FEEDBACK GENERATION

Feedback generation is a critical step in evaluating and refining songs. It involves comparing a song's features with reference data, identifying its strengths, and suggesting improvements. Here's an introduction and explanation of each concept within this step:

a. Compare with Reference :Comparing the extracted features of a song with reference features involves analyzing specific aspects of the song and benchmarking them against a database of well-rated songs. This helps in understanding how the song stands relative to successful compositions in the industry.

- **Feature Extraction:** Extract characteristics from the song, such as melody, harmony, rhythm, tempo, structure, and lyrics.
- **Reference Database:** A collection of features from well-rated songs that serve as a standard for comparison.
- **Benchmarking:** Evaluating how the song's features align with or differ from the reference data.

Process:

- Analyze the song to extract its musical features.
- Access the reference database containing features of well-rated songs.
- Compare each extracted feature with the corresponding reference feature to identify similarities and differences.

b. Identify Strengths : Identifying the strengths of a song involves recognizing the elements that stand out positively. This helps in acknowledging what works well in the song and can be emphasized further.

- **Positive Features:** Elements of the song that match or exceed the quality found in reference songs.
- **Unique Elements:** Aspects of the song that offer originality and creativity.

- **Performance Metrics:** Quantitative measures that indicate the song's strong points, such as high melody catchiness or strong lyrical content.

Process:

- Highlight features that closely align with or surpass reference features.
- Note unique elements that differentiate the song from others.
- Use performance metrics to objectively determine the song's strengths.

c. Suggest Improvements : Providing constructive feedback on areas that could be improved is essential for the refinement process. This step involves offering actionable suggestions based on the comparison and analysis.

- **Constructive Criticism:** Feedback that is aimed at helping the artist improve their work.
- **Improvement Areas:** Specific aspects of the song that do not meet the standard of reference songs or that could be enhanced.
- **Actionable Suggestions:** Practical advice and recommendations that the artist can implement to improve the song.

Process:

- Identify features that do not align well with reference data.
- Provide specific, actionable suggestions for improvement, such as altering the tempo, enhancing the lyrical content, or refining the song structure.
- Offer examples or techniques that can help in making the suggested improvements.

By systematically comparing the song with references, identifying its strengths, and suggesting improvements, feedback generation becomes a comprehensive process that guides artists in refining their work to meet higher standards and achieve greater success.

STEP 5: OUTPUT FEEDBACK

Providing effective feedback is a crucial part of any evaluation process, whether in education, workplace performance reviews, or customer feedback systems. Step 5 focuses on how to deliver this feedback effectively to ensure it is both useful and actionable.

a. Format Feedback: Formatting feedback in a clear, structured manner is essential for ensuring that the recipient understands the points being made. Poorly organized feedback can be confusing and counterproductive, potentially leading to misunderstandings or a lack of actionable insight.

- **Clarity:** Ensure that the language used is simple and straightforward. Avoid jargon unless the recipient is familiar with it.
- **Structure:** Use headings, bullet points, or numbered lists to organize feedback. This helps in breaking down the information into digestible parts.

- **Consistency:** Maintain a consistent format throughout to make it easier for the recipient to follow.
- **Brevity:** Be concise. Long-winded explanations can dilute the main points.
- **Examples:** Provide examples to illustrate points clearly. This can help in making abstract concepts more concrete.

b. Provide Recommendations: Simply pointing out areas that need improvement is not enough; it's crucial to offer specific, actionable recommendations. This ensures that the recipient knows exactly what steps to take to improve.

- **Specificity:** Avoid vague suggestions. Provide clear, specific advice that the recipient can act upon.
- **Actionable Steps:** Break down recommendations into manageable steps. This helps in creating a clear path to improvement.
- **Positivity:** Frame recommendations in a positive light. Instead of focusing solely on what was done wrong, highlight how things can be done better.
- **Context:** Provide context for the recommendations. Explain why certain changes are necessary and how they can lead to better outcomes.
- **Follow-Up:** Suggest a follow-up plan to review the implementation of the recommendations. This can help in maintaining accountability and ensuring continuous improvement.

CONCLUSION

In conclusion, this paper outlines a robust framework for an AI-based system designed to deliver insightful feedback on musical compositions. By integrating sophisticated techniques across preprocessing, feature extraction, and analysis phases, the algorithm effectively captures and evaluates various dimensions of a song, from its rhythmic and harmonic elements to its emotional and genre-specific characteristics. The structured feedback mechanism not only highlights a song's strengths but also provides actionable recommendations for improvement, making it a valuable tool for musicians and producers aiming to refine their work. This approach promises to enhance the quality of music production through data-driven insights and targeted suggestions.

REFERENCES

1. Librosa: Audio and Music Signal Analysis in Python" by Brian McFee et al.
2. Audio Resampling and its Effects on Music Information Retrieval" by Seyed Omid Sadjadi and John H.L. Hansen
3. A Tutorial on Audio Feature Extraction for Music Information Retrieval" by J. C. Brown
4. Music Feature Extraction and Processing: A Survey" by Peter Knees and Markus Schedl
5. Emotion Recognition from Speech and Music: A Survey" by Mohammad Soleymani et al.
6. A Survey on Music Genre Classification" by Thomas Lidy and Andreas Rauber
7. An Objective Metric for the Evaluation of Audio Quality" by C. M. O'Neill et al.

8. Content-based Music Recommendation: State-of-the-art, Challenges, and Future Directions" by E. Cano et al.
9. User-Centered Design and Evaluation of a Music Search Engine" by Markus Schedl et al.
10. Designing Usable Feedback Mechanisms for Interactive Machine Learning Applications" by Saleema Amershi et al.
11. Towards User-Aware Music Information Retrieval" by Albin E. Melin et al.