1. Introduction

Content-based technology has emerged from the development of multimedia signal processing and widespread of web applications. In recent years, audio, as an important media, has been gained more and more attention.

There are 2 types of features in audio, measurable acoustic and psychoacoustic features. Acoustic properties are how sound behaves within a given physical context like amplitude and loudness. Sound recognition and audiology properties of human knowledge are subjective terms perceived by the human auditory system and they include pitch, timbre, etc. It is natural to combine both points of view into the analysis. [4]

An audio is classified as speech, music, several types of environmental sounds, silence and many more. The audio database has details of feature extracted from various categories of audio. The proposed project allows a user to retrieve audio, based on the features of input audio request. An audio having same or similar features will be displayed to the user. [5]

The advanced libraries without bounds will incorporate (ASCII) content data as well as sounds alongside checked paper records and recordings. In this way, there is a need to recover data from such multi-media accumulations in light of its substance. This incorporates recovery of information identified with sound also. [9]

The techniques for information retrieval that exists currently are successful for text documents, whose evidence is given by the vast development of search engines like Google, Yahoo, and Bing in commercial profit. In comparison, for retrieval of multimedia data, there is no existing product or tool that offers satisfaction to clients as compared with the text-based search engines. Audio segmentation, speech recognition, audio information and environmental sound retrieval are areas of research in the content-based classification of audio data. [10]

By audio segmentation, we can distinguish several audios such as human voice, different melodies, silence and other environmental audios. It is mandatory to perform the pre-processing step which determines the similar parts of an audio stream. It also assists to further analyze the various audio types using the required techniques. [7]

Audio data recovery has turned into a critical zone in the most recent decade. It essentially manages recovery of comparative bits of music, instruments, specialists, melodic sorts, tones, human voice and the investigation of melodic structures. It additionally centers around pitch, assault, recurrence, length and flag wellspring of each solid in a bit of the sound record. [3]

1.1 Problem Definition

To design a methodology to classify and retrieve audio based on its content from the audio database.

1.2 Objectives

The main objective of content-based audio retrieval is the identification of similar audio data.

1. To design a methodology to retrieve audio based on the content of the requested audio:
    a. The entered audio should be of a .wav type.
    b. The entered audio should be of length 3 seconds to 10 seconds.
    c. The entered audio should have at least one similar audio file in the audio database.
2. There should be an audio file displayed:
    a. The output should have the most similar audio file from the audio database.
    b. There should be at least one similar audio file present in the database.
3. The listed audio file must match the input audio file as accurately as possible:
    a. The accuracy of matching of input audio file and the output audio file should be good enough.

2. Literature survey

Sounds are generally delineated by their pitch, clamor, timbre, and term. These insightful models are without a doubt knew and can be correctly shown by quantifiable acoustic features. A couple of researchers prototyped figuring prepared for expelling sound structure from a sound. These figuring were tuned to specific melodic creates and were not fitting for all sounds. Remarkable parts of tone quality consolidate the changes in the abundance of a sound after some time, vibration repeat, and adequacy plane. The ambush bits of a tone are routinely fundamental to perceive the timbre. [1]

Nevertheless, research has shown that the time-varying spectrum of a single musical instrument tone cannot generally be treated as a "fingerprint" identifying the instrument, because there is too much variation across the instrument's range of pitches and across its range of dynamic levels. [11]

Support vector machines have as of late proposed as another learning calculation for design acknowledgment. The support vector machines with a binary tree acknowledgment system/strategy is utilized to adapt the sound arrangement issue. They show the capability of support vector machines on a typical sound database. They look at the support vector machine-based grouping with other prevalent methodologies. [2]

There is a system named Muscle Fish for the same purpose. It is a famous content-based audio classification and retrieval software. Various perceptual features such as loudness, brightness, pitch, timbre, frequency, etc., are used to represent a sound, and the neural network rule is used for classification and retrieval of the audio data. It has a classification error rate as 19.07% as obtained from the Muscle Fish web interface. [8]

3. Methodology

We are implementing this project using python. Python is an interpreted high-level programming language for general-purpose programming. We are using audio database to store and retrieve audio files and its features.

The most important and challenging part of this project is the audio retrieval based on the content of the entered audio.

3.1 Description of proposed system

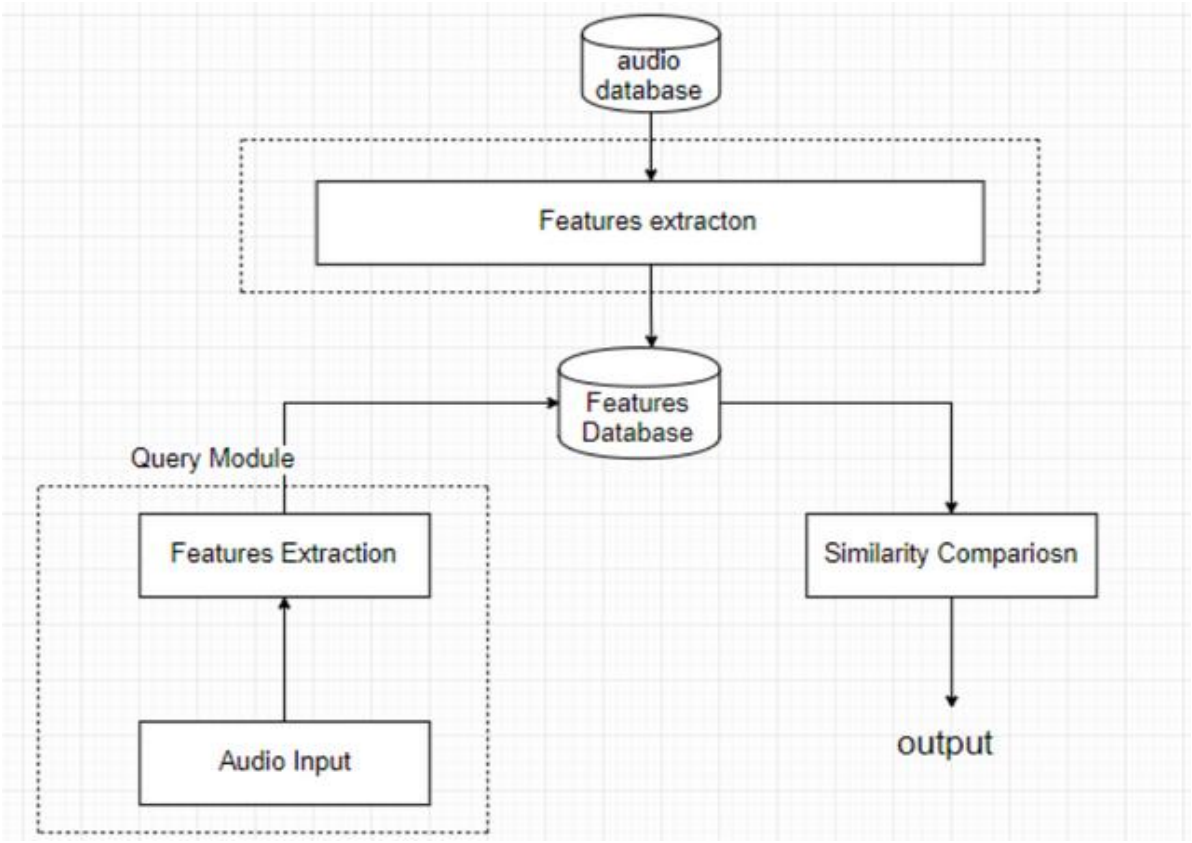

Figure 3.1 Architecture diagram of the proposed system

Figure 3.1 shows that the proposed methodology extracts the features from the audio database and gets trained. Further, when the user gives a different audio file as input, the audio file from the database with same or most similar features will be displayed as output.

The output will be based on:

1. Accurate match of audio samples.

2. Inaccurate match of audio, irrespective of sample rate, quantization, compression, etc.
3. Inaccurate match of features or perceptual properties of sound.

3.2 Requirement analysis

A requirement analysis is a description of a software system to be developed. It lays out functional and non-functional requirements, and may include a set of use cases that describe user interactions that the software must provide.

3.2.1 Functional requirements

1. The system shall be able to extract features.
2. The system shall be able to compare similar features.
3. The system shall be able to retrieve audio features from the database through queries.
4. The system shall be able to classify audios based on the stored features.

3.2.2 Non-functional requirements

1. Maintainability: The software shall be easy to maintain and develop.
2. User-friendly: The user shall be able to use it without any difficulties.
3. Reliability: The output shall always be 85-95% efficient.

3.3 Design

Systems design is the process of defining the architecture, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development. In our project as a part of system design we have done the detailed design of the system.

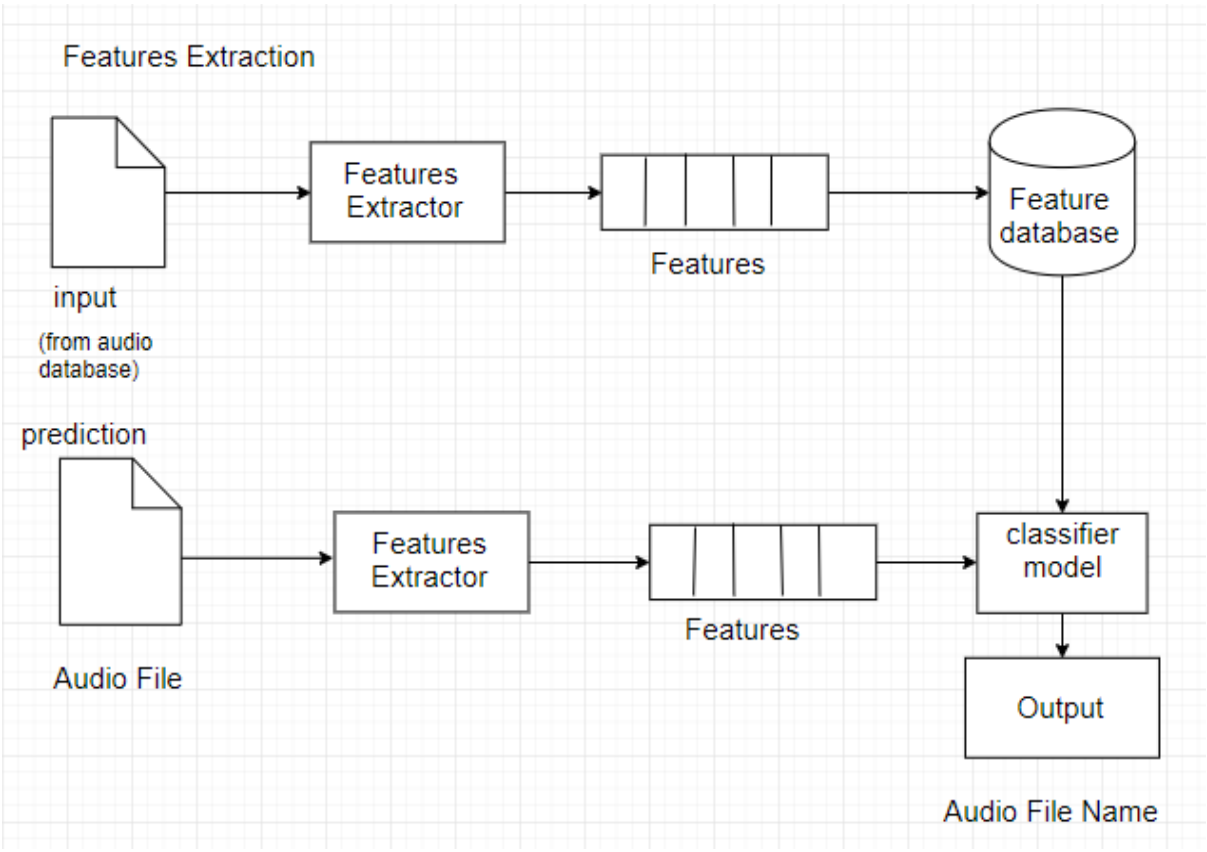

Figure 3.2 Detailed design of the proposed system

Figure 3.2 shows the detailed design of content base audio retrieval. The proposed software extracts the features from the audio database given as input to the software and gets trained and when the user gives an audio file as input it lists down all the audio files with similar features as output.

Here, the features extracted include:

1. Energy and RMSE: A frequently used measure of the differences between values.
2. Zero crossing rate: This is the rate at which the signal converts from positive to negative or negative to positive.
3. Magnitude scaling (amplitude): The highest extent of an oscillation or a vibration, measured from the position of equilibrium.
4. Pitch transcription (onsets): The fall and rise of a tone.
5. Tempo (global): The speed at which a passage of music is or should be played.
6. Beat tracking: The detection of a beat in an audio.

4. Limitations

There are a few limitations in this project. They're listed as follows:

1. The length of the audio file to be given as input should not exceed the limit of 3-10 seconds.
2. The audio must be of the type .wav file.

5. Results and Discussions

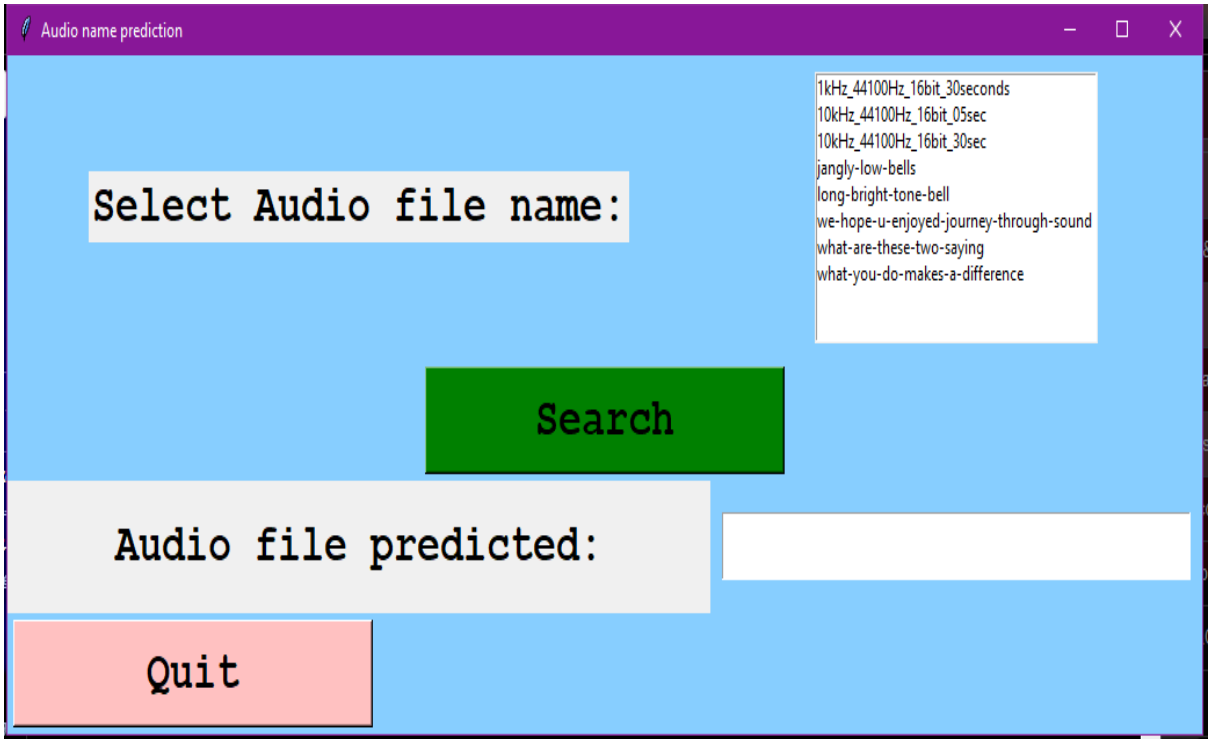

Figure 5.1 Fetching of the audio file

Figure 5.1 shows the graphical user interface of the software where it takes an audio file as input and searches for the most similar audio.

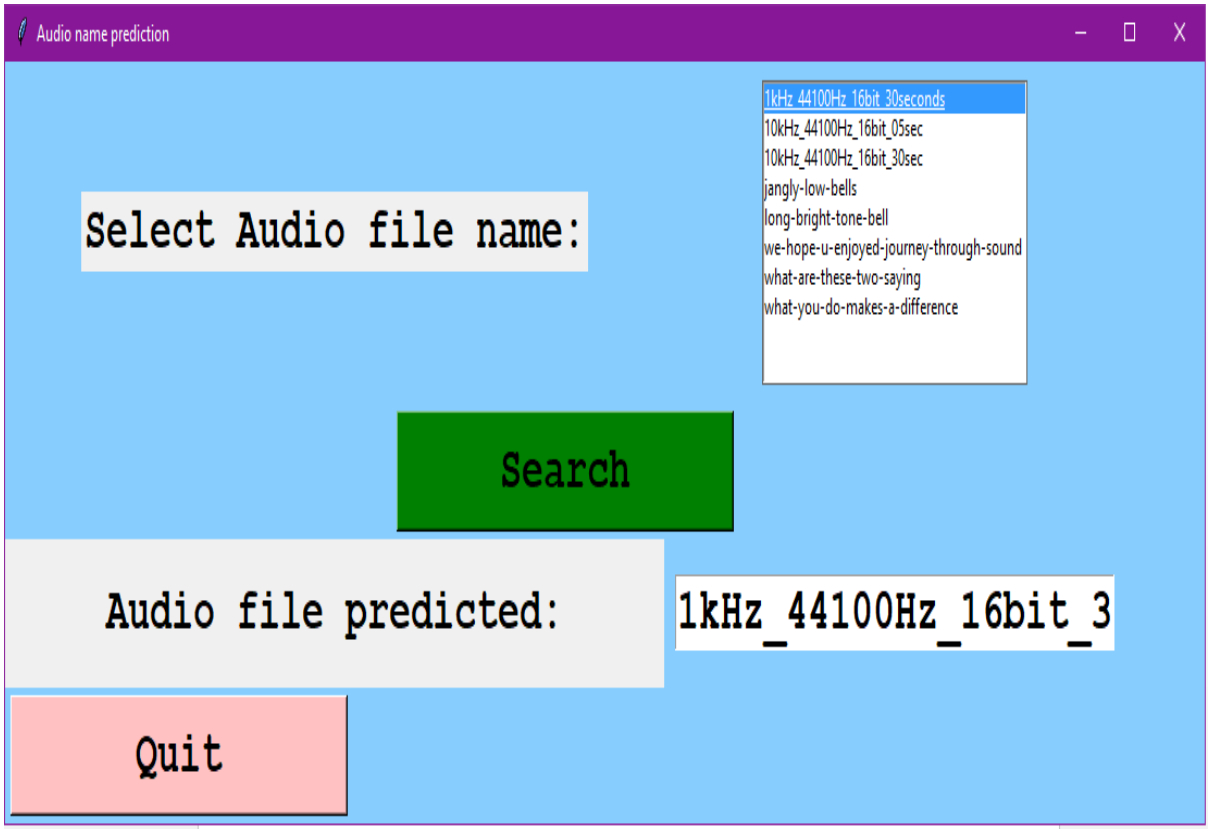

Figure 5.2 Most similar audio from database as output

Figure 5.2 shows the graphical user interface of the output. It is displaying the most similar audio file from the database.

6. Conclusions

In this software, when an audio file is given as input, an audio file from the trained dataset is retrieved. The exact or most similar audio file will be retrieved. Our approach focuses on analyzing six different features. These distinct features involve energy and RMSE, zero crossing rate, magnitude scaling (amplitude), pitch transcription (onsets), tempo (global) and beat tracking of .wav audio files and realize the query-by-example music retrieval in a database. As the audio files are limited to a length of 3-10 seconds and of type .wav, we can extend this project to longer audio length and to mp3 format.

7. References

[1] Erling wold, Thom Blum, Douglas Keisler, and James Wheaton, “content-based classification, search and retrieval of audio” IEEE transactions on neural networks, vol. 14, no. 1, January 2003.

[2] Guodong Guo and stan z. Li, “content-based audio classification and retrieval by support vector machines”.

[3] Nilesh M. Patil, Milind U. Nemade, “Content-based audio classification and retrieval: A novel approach” Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 2016.

[4] Stan Z. Li, “Content-based audio classification and retrieval using the nearest feature line method”, IEEE Transactions on Speech and Audio Processing, Volume: 8, Issue: 5, Sep 2000.

[5] R. Christopher, Praveen Kumar, S. Suguna, J. Becky Elfreda, "Audio Retrieval based on Cepstral Feature", International Journal of Computer Applications, vol. 107, no. 17, pp. 28-33, December 2014, ISSN 0975-8887.

[6] A. Yoshitaka, T. Ichikawa, “A survey on content-based retrieval for multimedia databases”, IEEE Transactions on Knowledge and Data Engineering, vol. 5, no. 4, pp. 619-628, 1993.

[7] Jonathan T. Foote, “Content-based retrieval of music and audio”, Multimedia Storage and Archiving Systems- II, 6 October 1997.

[8] Guohui Li, A.A. Khokhar, “Content-based indexing and retrieval of audio data using wavelets”, 2000 IEEE International Conference on Multimedia and Expo. ICME2000. Proceedings. Latest Advances in the Fast-Changing World of Multimedia (Cat. No.00TH8532), July 30, 2000-Aug. 2 2000.

[9] Michael A. Casey, Remco Veltkamp, Masataka Goto, “Content-Based Music Information Retrieval: Current Directions and Future Challenges”, Proceedings of the IEEE, Volume 96 Issue 4, April 2008.

[10] Ishwar K. Sethi, Nevenka Dimitrova, Tom McGee, “Classification of general audio data for content-based retrieval”, Pattern recognition letters, Volume 22, Issue 5, April 2005, Pages 533-544.

[11] Mingchun Liu, Chunru Wan, “A study on content-based classification and retrieval of audio database”, Proceedings 2001 International Database Engineering and Applications Symposium, 16 – 18 July 2001.