MUSICAL FREQUENCY NOTE DETECTION

Kaustubh Joshi[1], Manish Godbole[2], Aditya Kadu[3], Rekha Kulkarni[4], Mukta Takalikar[5]

1 Pune Institute of Computer Technology, Computer Engineering, Pune, Maharashtra, India, knj621311@gmail.com
2 Pune Institute of Computer Technology, Computer Engineering, Pune, Maharashtra, India, manishgodbole02@gmail.com
3 Pune Institute of Computer Technology, Computer Engineering, Pune, Maharashtra, India, adityakadu1203@gmail.com
4 Pune Institute of Computer Technology, Computer Engineering, Pune, Maharashtra, India, rakulkarni@pict.edu
5 Pune Institute of Computer Technology, Computer Engineering, Pune, Maharashtra, India, mstakalikar@pict.edu

Abstract

Accurate and rapid detection of musical notes is essential for tasks such as automatic tuning, transcription, and instrument recognition. Our intended work employs detector which uses advanced signal processing techniques to analyze audio input and determine the fundamental frequency (pitch) of the predominant musical note being played.

The system utilizes a combination of time-domain and frequency-domain analysis to extract relevant features from the input audio signal. These features are then fed into a machine learning-based classifier that identifies the closest musical note corresponding to the detected frequency. To ensure adoptability and accuracy, the system has been trained on a comprehensive dataset covering a wide range of musical instruments, playing styles. Our work presents the design, implementation, and evaluation of a novel musical frequency note detector aimed at instrumental applications in varied musical contexts.

Keywords: *Audio input, Fundamental frequency, Pitch, Time-domain, Frequency-domain, Novel musical frequency note detector.*

I. INTRODUCTION

Music, a global medium that bridges geographical and cultural divides, has consistently been a focus of intrigue and research. The capacity to dissect and comprehend music's subtleties, such as identifying individual musical notes within a piece, has extensive implications in areas from music theory and pedagogy to audio processing and digital signal examination. This research article ventures into the sphere of musical frequency note detection, a vital field of study in music technology. By investigating the principles and methodologies underlying this process, we aspire to illuminate the fundamental mechanics of musical notes, their frequencies, and how cutting-edge technology can aid their accurate identification. In doing so, we anticipate opening of new pathways for creativity, education, and innovation in the music domain.

II. LITERATURE REVIEW

Akhilesh Sharma, et al. [1], have studied about music information retrieval, focusing on Indian Classical music and its two major parts. Authors have used classification methods like Mel frequency cepstral coefficients (MFCCs) and spectrograms.

It considers the computational techniques applied to understand the heritage of Indian Classical Music, highlighting the distinct characteristics of Hindustani and Carnatic traditions. It explores the complexity of Carnatic Music, emphasizing the significance of Ragas and Talams. The paper concludes by examining the fundamental elements of Indian Classical Music, such as musical notes, intervals, and scales, highlighting the unique aspects of Hindustani and Carnatic traditions.

Artificial neural networks have witnessed three notable waves: the perceptron algorithm (1957), backpropagation algorithm (1986), and the deep learning success in 2012. Hendrik Purwins et al. [2], have extensively studied Deep learning networks including architectures like deep feedforward neural networks, convolutional neural networks (CNNs), and long short-term memory (LSTM). Due to a focus on audio signal processing, this wave has surpassed traditional methods, particularly in image, speech, music, and environmental sound processing. While looking at image processing, audio has unique challenges due to its one-dimensional time series nature. The shift to deep learning in several domains, has outperformed conventional models where ample data is available.

As per the work by Jay K. Patela et al. [6], A song contains basically two things, vocal and background music. Where the characteristics of the voice depend on the singer and in case of background music, it involves mixture of different musical instruments like piano, guitar, drum, etc. To extract the characteristic of a song becomes more important for various objectives like learning, teaching, composing. The experiment is done with the several piano songs where the notes are already known, and identified notes are compared with original notes until the detection rate goes higher. And then the experiment is done with piano songs with unknown notes with the proposed algorithm.
The article by John Glover et al. [7] provides a review of some of the most used techniques for real-time onset detection. The authors suggest ways to improve these techniques by incorporating linear prediction as well as presenting a novel algorithm for real-time onset detection using sinusoidal modelling. As well as provides comprehensive results for both

the detection accuracy and the computational performance of all the described techniques, evaluated using Modal.

In the research by Allabakash Isak Tamboli et al. [8], the authors developed a musical note recognition method based on an optimization-based neural network (OBNN) within a classification framework. The study involved an extensive review of existing approaches for musical note recognition. The use of OBNN for recognizing musical notes was explored. The document comprehensively analyzes recent investigations related to musical note recognition, summarizing their findings and classifications, with the aim of advancing the effectiveness of this recognition process through diverse methodologies.

The paper by Smith Julius O. [3] gives seminal work in the field of digital audio processing. This paper delves into the principles and methodologies of physical modeling, which simulates the behavior of real-world musical instruments and sound effects in the digital domain. It explores the mathematical and computational foundations of physical modeling, allowing for the creation of highly realistic virtual instruments and audio effects. By emphasizing the accurate emulation of physical interactions and acoustic phenomena, Smith's research paper has been pivotal in advancing the quality and authenticity of digital music synthesis and audio processing. It remains a foundational reference for researchers and engineers in the field.

The paper by H Purwins et al. [2]. gives comprehensive overview of the application of deep learning techniques in the field of audio signal processing. It explores the use of neural networks and deep learning architectures for tasks such as speech recognition, music analysis, and sound synthesis. The paper discusses various deep learning models and their effectiveness in handling complex audio data. It serves as a valuable resource for researchers and practitioners interested in leveraging deep learning for advanced audio processing applications.

The authors of [4] present Critical problem of accurately estimating pitch in speech signals contaminated by noise. The authors propose a novel pitch estimation method tailored for noisy conditions, focusing on the challenging scenario of adverse environmental or recording conditions. Their approach combines adaptive filtering and signal processing techniques to enhance the accuracy and robustness of pitch estimation in the presence of noise. This paper presents an essential contribution to speech signal processing, particularly in contexts where noise interference poses a significant challenge, making it valuable for applications like speech recognition and enhancement.

The work by S. Wang et al. [5] senses self-supervised learning approach that leverages audio-visual data with spatial alignment to enhance audio representation learning. The proposed method combines visual information and audio signals to train deep neural networks without explicit annotations. By exploiting spatial alignment cues, the model learns robust and informative representations, which have applications in areas such as speech and sound analysis, offering potential benefits for improving the accuracy of audio-based tasks using multi-modal data.

III. METHODOLOGY

In our proposed method we are using the combination of signal processing, and visualization for the analysis and comprehension of audio properties. Amplitude envelope is calculated to know about the peak amplitude within chosen frame sizes. To align the envelope with time for further visualization time and the frame calculations are done. The Short-Term Fourier Transform (STFT) is used to find time frequency characteristics of the audio. Pitch identification is done by tracking the peaks in the magnitude of STFT. Finally, the note mapping method is used in mapping frequencies to musical note. Following methods have been used as a part of our implementation.

1. **Audio Loading**: Audio Import: The code initiates by importing an audio file ('test.mp3') utilizing the librosa.load function, which results in the raw audio waveform and its sampling rate (SR). This action readies the data for further examination.

2. **Waveform Visualization**: It continues with the display of the audio waveform using Matplotlib. This display illustrates the amplitude of the audio signal as it progresses over time, offering a visual comprehension of the audio's attributes.

3. **Amplitude Envelope**: To scrutinize the signal's fluctuations, two functions, amp_env and fancy_amp, are employed to calculate the amplitude envelope. The amplitude envelope seizes the peak amplitude within designated frame sizes, a vital feature for a variety of audio processing tasks.

4. **Time and Frame Calculation**: The code begins by determining the time and frame indices for the amplitude envelope using the librosa.frames_to_time function. This step aligns the envelope with time for subsequent visualization.

5. **Visualizing the Envelope**: The code then generates another Matplotlib plot, which displays the audio waveform and overlays the amplitude envelope in red. This visualization aids in understanding the amplitude's temporal variations.

6. **Short-Time Fourier Transform (STFT)**: To delve into the audio's time-frequency characteristics, the code computes the STFT using `librosa.stft`. The STFT provides a detailed representation of the audio signal in the time and frequency domains.

7. **Pitch Detection**: For pitch analysis, the code uses the `librosa.piptrack` function to identify pitch frequencies in each frame. This process is achieved by tracking the peaks in the magnitude of the STFT.

8. **Note Mapping**: A dictionary, `note_mapping`, is defined to map detected frequencies to their corresponding musical note names, allowing for easier interpretation of the pitch information.

9. **Average Pitch Calculation**: The code calculates the average pitch within specific frame intervals to provide a more generalized view of the audio's pitch characteristics. It calculates the mean pitch, ignoring any NaN values in the pitch data.

10. **Displaying Results**: Finally, the code prints and displays the average frequency and corresponding musical notes for each set of frames at specified intervals.

Our system architecture diagram is depicted in the following Fig. 1

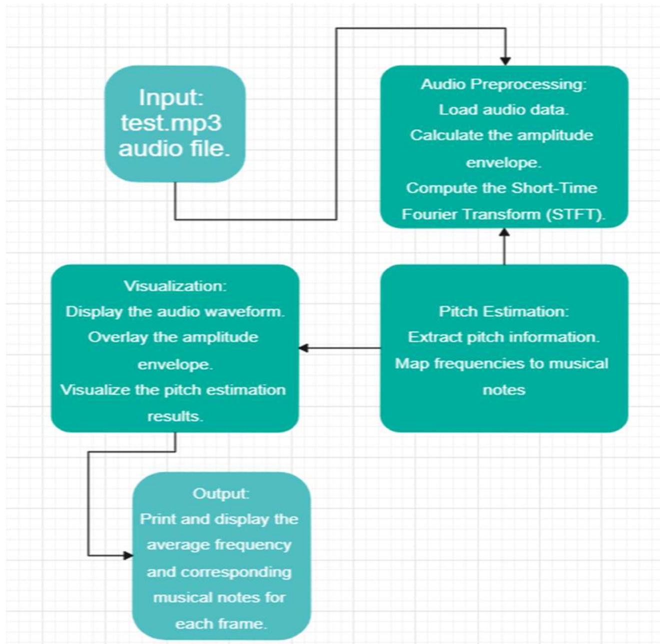

Fig 1 Proposed Architecture System

IV. IMPLEMENTATION

We integrate signal processing, visualization, and feature extraction techniques to analyze audio data, highlighting attributes such as amplitude variations and pitch characteristics in a structured and visually informative manner.

The methodology primarily relies on traditional audio analysis techniques and visualization, rather than machine learning models or classifiers.

V. RESULTS AND DISCUSSION

We have implemented our method for finding the musical notes for which following table shows some sample results.

Table 1: Sample Frequency and Corresponding Note

```
Average Frequency and Notes for Every 132 Frames:
Frames 1-132: 549.40 Hz, None
Frames 133-264: 557.13 Hz, None
Frames 265-396: 523.65 Hz, None
Frames 397-528: 434.92 Hz, A4
Frames 529-660: 445.90 Hz, A4
Frames 661-792: 393.01 Hz, G4
Frames 793-924: 409.39 Hz, G#4
Frames 925-1056: 377.58 Hz, F#4
Frames 1057-1188: 374.47 Hz, F#4
Frames 1189-1320: 298.26 Hz, D4
Frames 1321-1452: 400.85 Hz, G4
Frames 1453-1584: 452.31 Hz, None
Frames 1585-1716: 419.64 Hz, G#4
Frames 1717-1848: 312.35 Hz, D#4
Frames 1849-1980: 331.72 Hz, E4
Frames 1981-2112: 322.05 Hz, E4
Frames 2113-2244: 427.73 Hz, None
Frames 2245-2376: 337.00 Hz, E4
Frames 2377-2508: 313.32 Hz, D#4
Frames 2509-2640: 424.86 Hz, G#4
Frames 2641-2772: 317.97 Hz, D#4
Frames 2773-2904: 423.64 Hz, G#4
Frames 2905-3036: 367.24 Hz, F#4
Frames 3037-3168: 274.59 Hz, C#4
Frames 3169-3300: 347.83 Hz, F4
Frames 3301-3432: 574.12 Hz, None
Frames 3433-3564: 529.09 Hz, None
Frames 3565-3696: 502.88 Hz, B4
Frames 3697-3828: 430.95 Hz, A4
Frames 3829-3960: 429.46 Hz, None
Frames 3961-4092: 391.47 Hz, G4
```

Fast Fourier Transform (FFT) is a mathematical technique that transforms an audio signal into its frequency components. It's fast and efficient but requires a lot of memory and may not be suitable for real-time applications.

Librosa is a Python package for music and audio analysis. It's easy to use and offers a rich set of features for music analysis. However, it may not be compatible with some older versions of Python or operating systems.

Autocorrelation is a statistical measure that quantifies the similarity between a given time series and its lagged version. It can reveal hidden information in data but requires a lot of computation time and may produce misleading results if the data is non-stationary or has noise.

METHOD	PARAMETER	BEST APPLICATION
FFT	Frequency Resolution	Audio Signal Processing
LIBROSA	MFCC	Music Genre Classification
AUTOCORRELATION	Pitch Period	Speech Processing

APPLICATION TABLE

VI. CONCLUSION

The primary objective of this study is to comprehend the detection of musical frequency notes, highlighting the immense possibilities for application development and its significant influence on music and technology. The necessity for precise note detection spans various areas, including music education, transcription, and audio processing. Accurate note identification provides essential tools for musicians and learners, enhancing music teaching and learning, and allowing musicians to perfect their performances and compositions. Moreover, our research has demonstrated how advancements in technology, such as digital signal processing and machine learning, have streamlined and democratized automated note detection, paving the way for the creation of software tools and applications that assist musicians of all skill levels. In addition, our research emphasizes the critical need for meticulous investigation in this field, underscoring the importance of extensive and diverse datasets, sturdy algorithms, and the ongoing refinement of techniques to boost the precision and dependability of note detection systems. In essence, the goal of musical frequency note detection extends beyond the technical sphere, fusing creativity and education with technology. This harmonious integration of music and technology opens new avenues for articulating artistic work with learning. As we persist in refining and innovating, we envision a future where music becomes more accessible, lucid, simple, and enriched for everyone. This research uncovers the immense potential that awaits in the domain of musical note detection.

REFERENCES

1. Akhilesh Sharma, Gaurav Aggarwal, Sachit Bharadwaj, Prasun Chakrabarti, Tulika Chakrabarti, Jemal H. Abawajy, Siddhartha Bhattacharyya, Richa Mishra, Anirban Das, Hairulnizam Mahdin, "Classification of Indian Classical Music with Time-Series Matching Deep Learning Approach", IEEE Access, vol. 9.

2. Hendrik Purwins , Bo Li , Tuomas Virtanen , Jan Schluter ¨ , Shuo-Yiin Chang, and Tara Sainath, "Deep Learning for Audio Signal Processing" in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, May 2019.

3. Smith, Julius O. "Physical Audio Signal Processing for Virtual Musical Instruments and Audio Effects." W3K Publishing, 2010.

4. S. A. Shedied, M. E. Gadalah and H. F. VanLundingham, "Pitch estimator for noisy speech signals," Smc 2000 conference proceedings. 2000 ieee international conference on systems, man and cybernetics. 'Cybernetics evolving to systems, humans, organizations, and their complex interactions' (cat. no.0, Nashville, TN, USA, 2000, pp. 97-100 vol.1, doi: 10.1109/ICSMC.2000.884971.

5. S. Wang, A. Politis, A. Mesaros and T. Virtanen, "Self-Supervised Learning of Audio Representations from Audio-Visual Data Using Spatial Alignment," in IEEE Journal of Selected Topics in Signal Processing, vol. 16, no. 6, pp. 1467-1479, Oct. 2022, doi: 10.1109/JSTSP.2022.3180592.

6. Jay K. Patela, E.S.Gopia. "Musical Notes Identification using Digital Signal Processing". 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015)

7. John Glover, Victor Lazzarini and Joseph Timoney. "Real-time detection of musical onsets with linear prediction and sinusoidal modeling". EURASIP Journal on Advances in Signal Processing.

8. Allabakash Isak Tamboli and Rajendra D. Kokate. "An Effective Optimization-Based Neural Network for Musical Note Recognition".