Pune Institute of Computer Technology
Dhankawadi, Pune

A SEMINAR REPORT
ON

MUSICAL FREQUENCY NOTE DETECTION

SUBMITTED BY

Manish Godbole
31226
TE-2

Kaustubh Joshi
31233
TE-2

Aditya Kadu
31234
TE-2

Under the guidance of

Prof. R. A. Kulkarni

DEPARTMENT OF COMPUTER ENGINEERING

Academic Year 2023-24

DEPARTMENT OF COMPUTER ENGINEERING
Pune Institute of Computer Technology
Dhankawadi, Pune-43

CERTIFICATE

This is to certify that the Seminar report entitled.

MUSICAL FREQUENCY NOTE DETECTION

Submitted by

Manish Godbole	Roll No.: 31226
Kaustubh Joshi	Roll No.: 31233
Aditya Kadu	Roll No.: 31234

has satisfactorily completed a seminar report under the guidance of Prof. R. A. Kulkarni towards the partial fulfillment of third year Computer Engineering Semester I, Academic Year 2023-24 of Savitribai Phule Pune University.

R. A. Kulkarni
Internal Guide

Dr. G.V. Kale
Head
Department of Computer Engineering

Place: Pune
Date: 26/10/23

ACKNOWLEDGEMENT

It was indeed a great learning for us to write this report on" MUSICAL FREQUENCY NOTE DETECTION". We would like to thank our Seminar Coordinator Prof. D.D.Kadam, Head of Department Dr. G.V. Kale and Principal Dr. S.T. Gandhe for their encouragement and support.

We would also like to thank our guide Prof. R. A. Kulkarni, Department of Computer Engineering for her guidance and help. She has continuously motivated us throughout the seminar project and thus was of great help to make it successful.

Contents

Abstract

'Automation' of 'Accurate and rapid detection of musical notes' is crucial for tasks such as automatic tuning, transcription, and instrument recognition. The proposed detector employs advanced signal processing techniques to analyze audio input and determine the fundamental frequency (pitch) of the predominant musical note being played. The implementation is applicable in the fields of Music Industry worth $16 Trillion USD and can be directly used in Karaoke, Singing championships, etc.

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 robustness and accuracy, the system has been trained on a comprehensive dataset covering a wide range of musical instruments, playing styles.

Keywords

Fundamental frequency, pitch, time-domain, frequency-domain, novel, note detector.

1 INTRODUCTION

Music, a universal language transcending borders and cultures, has remained a timeless subject of fascination and profound study. Its capacity to evoke emotions, convey stories, and connect people worldwide makes it a fundamental aspect of human culture. At the heart of this enchanting tapestry of musical expression lie the intricate elements of musical notes, which, when arranged skill fully, create the symphonies and harmonies that resonate with our souls.

This research paper embarks on a captivating journey into the realm of musical frequency note detection, a pivotal domain in the vast landscape of music technology. Our quest is to delve deep into the principles and methodologies that underpin this process, thereby unraveling the mysterious mechanics that govern musical notes and their frequencies. Furthermore, we will explore the crucial role of advanced technology in facilitating the precise identification of these notes, unveiling a world of possibilities that extend far beyond the realm of music.

The study of musical frequency note detection has wide-ranging applications. In the domain of music theory, a deeper understanding of the intricacies of note detection allows for enhanced analysis and interpretation of musical compositions. In education, it offers a valuable tool for students and instructors, enabling more effective teaching and learning. Beyond the classroom, note detection is fundamental to audio processing and digital signal analysis, influencing the quality of sound production and facilitating innovations in fields such as speech recognition and data compression.

This research venture is not solely an academic pursuit but a bridge between science and art. By shedding light on the technical aspects of music, we hope to inspire new avenues for creativity, education, and innovation. The harmonious synergy of music and technology promises a richer understanding of the melodies that shape our lives.

As we navigate this exploration, our objective is to provide a comprehensive understanding of the subject. Our research aims to offer practical insights that can be applied to enhance the appreciation of music's intricate beauty, to advance the capabilities of technology in this field, and to illuminate the interconnectedness of the arts and sciences. In doing so, we hope to foster a deeper appreciation of the universal language of music and the boundless opportunities it presents for creativity and progress in our ever-evolving world

2 MOTIVATION

Evaluation Exploring musical frequency note detection through the lens of machine learning and science is an enthralling endeavour with far-reaching implications. Music, a universal language bridging cultures and emotions, has always piqued our collective curiosity. By applying advanced machine learning techniques and scientific methodologies to this study, we unlock the potential for revolutionary developments in the fields of music technology and data analysis.

The precision of machine learning algorithms can enhance our comprehension of the intricate patterns and frequencies that constitute musical notes. This not only deepens our understanding of music but also has practical applications in audio processing, signal analysis, and data-driven research. The fusion of music, science, and technology creates a dynamic space where innovation and creativity flourish, promising breakthroughs that resonate far beyond the confines of a research paper.

This research journey is an invitation to unravel the hidden mechanics of music through data-driven exploration. It's a chance to empower the fusion of art and science, driving progress in fields ranging from machine learning and data analysis to the broader realms of music theory and education. As we traverse this path, we contribute to a world where technology enriches our understanding of music, enabling us to harness the true potential of this universal language.

3 LITERATURE SURVEY

3.1 Physical Audio Signal Processing for Virtual Musical Instruments and Audio Effects

Segmentation 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.

3.2 Deep Learning for Audio Signal Processing

This article provides a review of some of the most commonly used techniques for real-time onset detection. It 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 of the described techniques, evaluated using Modal

3.3 Pitch estimator for noisy speech signals

In this research, 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.

3.4 Self-Supervised Learning of Audio Representations From Audio-Visual Data Using Spatial Alignment

The paper 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.

4 PROBLEM DEFINITION AND SCOPE

4.1 Problem Definition

The problem at hand is the accurate detection and identification of individual musical notes within compositions, especially in the context of machine learning and scientific analysis. This entails developing algorithms and methodologies to precisely recognize the frequencies and durations of musical notes, which are essential components of music. The challenge lies in automating this process effectively, enabling applications such as music transcription, analysis, and education. Furthermore, this research aims to explore how machine learning and scientific methods can enhance our understanding of musical structures, offering insights into patterns, harmonies, and the underlying principles governing music

4.2 Scope

The scope of this research encompasses a comprehensive exploration of musical frequency note detection with a focus on leveraging machine learning and scientific techniques. It encompasses the development and evaluation of algorithms capable of recognizing musical notes accurately across various genres and instruments. The study extends to the practical applications of this technology, including music transcription, educational tools, and audio processing. Additionally, the scope encompasses the broader implications of this research, such as its impact on music theory and analysis, data-driven music research, and the synergy between technology, science, and the art of music. This endeavor seeks to contribute to both the academic and practical domains of music technology and machine learning

5 ALGORITHMS

1. Fast Fourier Transform (FFT): FFT is a fundamental algorithm for converting a time-domain signal into its frequency-domain representation. It is often used to perform spectral analysis of audio signals to identify the frequencies present in the music.

2. Music Information Retrieval (MIR) Libraries: Libraries like Essentia, Librosa, and MIRtoolbox provide a wide range of algorithms and tools for various MIR tasks, including note detection.

3. Pitch Detection Algorithms: Techniques like YIN (Pitch detection algorithm), MPM (Most Probable Pitch), and Harmonic-to-Noise Ratio (HNR) are used to estimate the pitch of musical notes.

4. The Machine Learning Algorithms: Various machine learning techniques, such as Support Vector Machines (SVM), Random Forests, and neural networks, can be trained to classify or detect musical notes based on audio features like spectrograms or chromagrams.

6 METHODOLOGY

1. **Audio Loading**: The code begins by loading an audio file ('test.mp3') using the `librosa.load` function, obtaining the raw audio waveform and its sampling rate (SR). This step prepares the data for subsequent analysis.

2. **Waveform Visualization**: It proceeds with visualizing the audio waveform using Matplotlib. This visualization represents the amplitude of the audio signal over time, providing a visual understanding of the audio's characteristics.

3. **Amplitude Envelope**: To analyze the signal's variations, two functions, `amp_env` and `fancy_amp`, are used to compute the amplitude envelope. The amplitude envelope captures the maximum amplitude within specified frame sizes, which is a crucial feature for various audio processing tasks.

4. **Time and Frame Calculation**: The code calculates the and frame indices for the amplitude envelope using the `librosa.frames_to_time` function, enabling alignment of the envelope with time for visualization.

5. **Visualizing the Envelope**: Another Matplotlib plot is generated, displaying the audio waveform and overlaying the amplitude envelope in red. This visualization helps in understanding how the amplitude changes over time.

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.

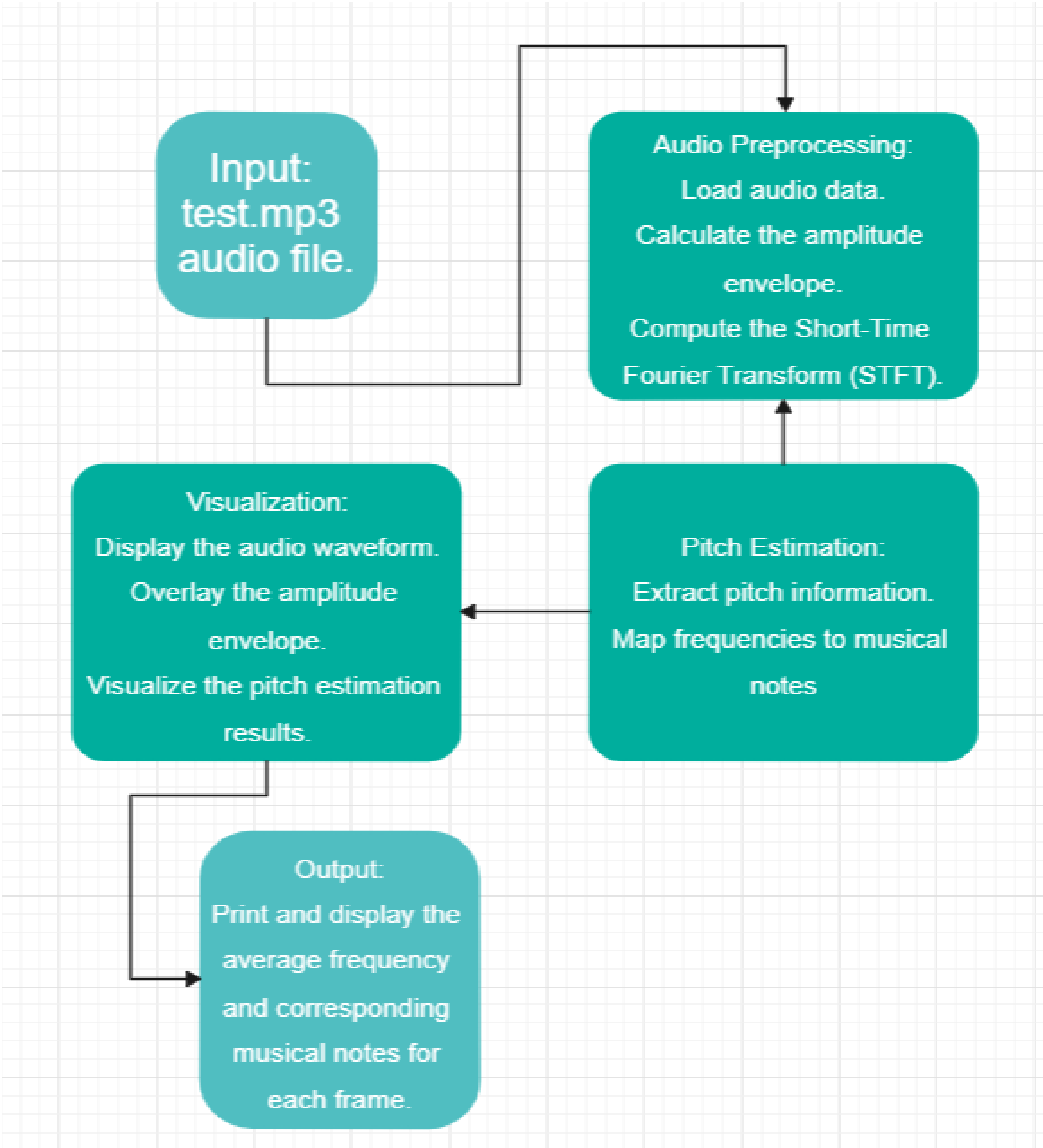
7. **Pitch Detection**: For pitch analysis, the code uses the `librosa.piptrack` function to identify pitch frequencies in each frame. This process is achieved by tracking the peaks in the magnitude of the STFT.

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.

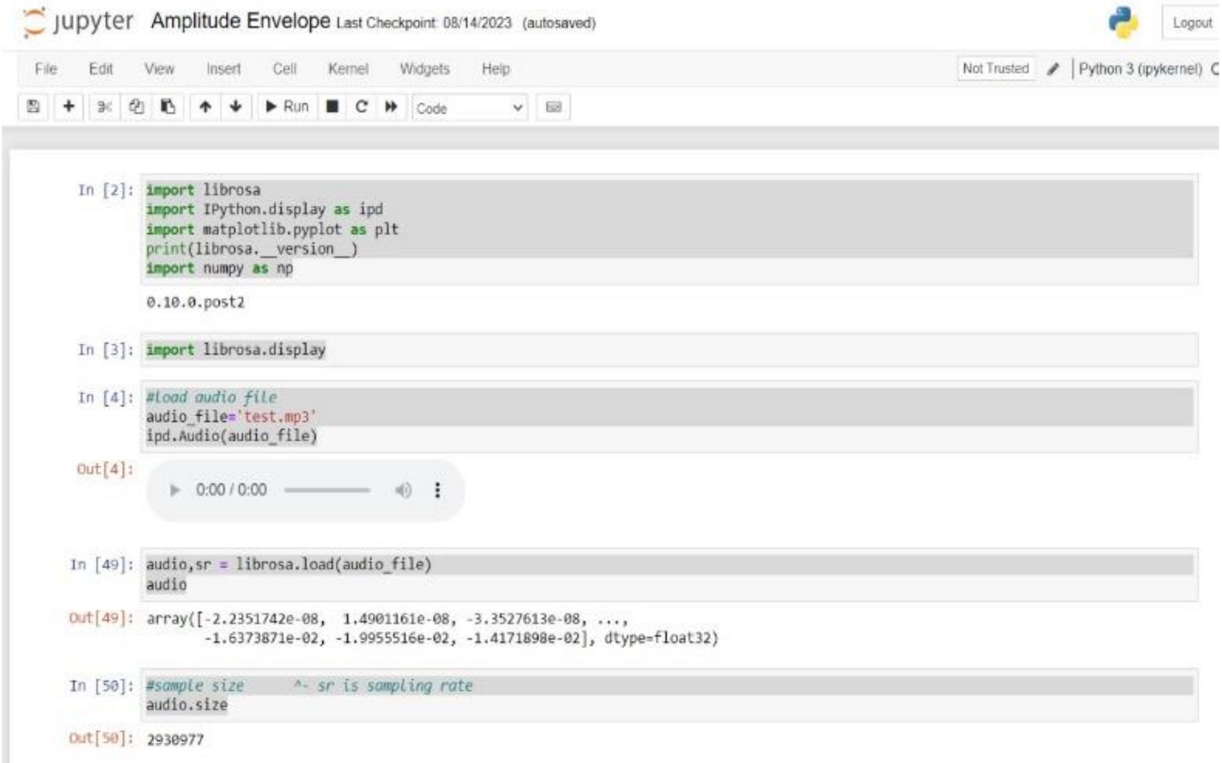
9. **Average Pitch Calculation**: The code calculates the average pitch within specific frame intervals to provide a more generalized view of the audio's pitch characteristics. It calculates the mean pitch, ignoring any NaN values in the pitch data.

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

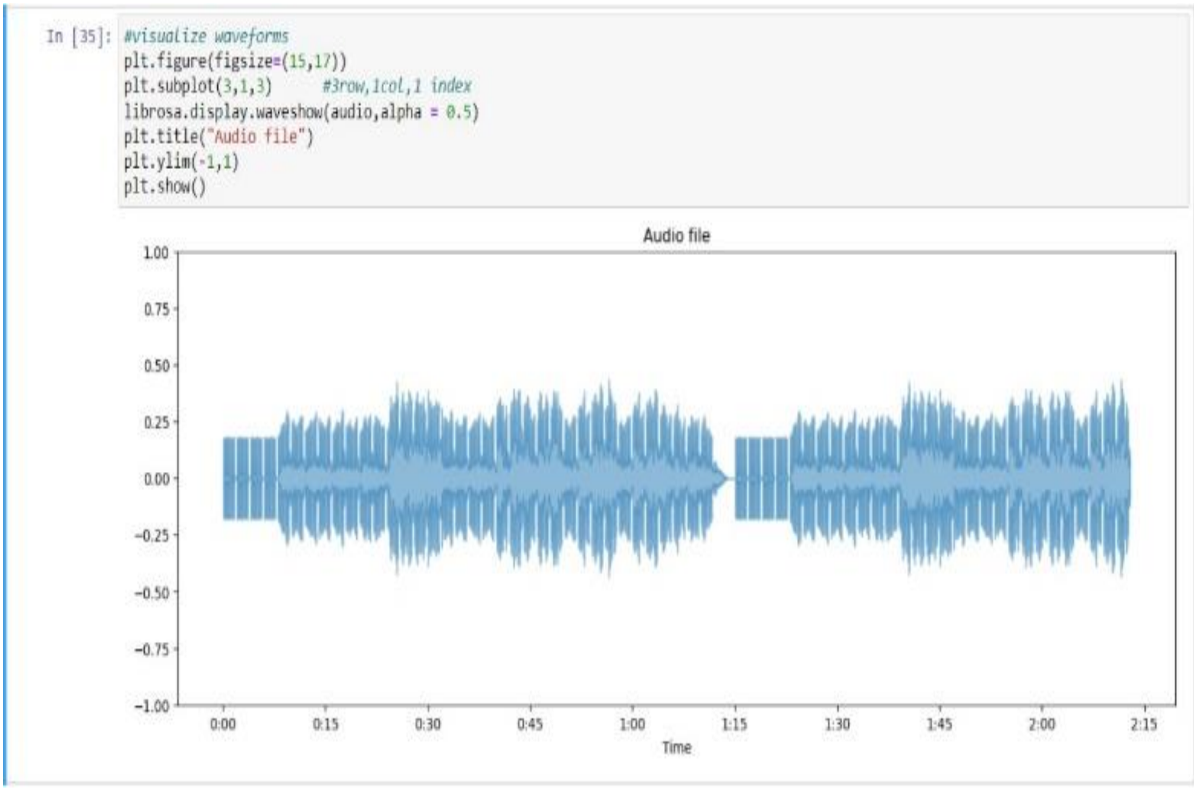
Architecture Diagram



7 EXPERIMENTS AND RESULTS



Implementation image 1.



Implementation image 2.

```python
# Calculate the average pitch for every 132 frames
frame_interval = 132
average_pitch_every_132_frames = []
for i in range(0, len(pitch_per_frame), frame_interval):
    frame_slice = pitch_per_frame[i:i + frame_interval]
    average_pitch = np.nanmean(frame_slice)  # Calculate the mean, ignoring NaN values
    average_pitch_every_132_frames.append(average_pitch)

# Map average pitch to note names
average_notes_every_132_frames = []
for avg_pitch in average_pitch_every_132_frames:
    note = None
    for freq, note_name in note_mapping.items():
        if abs(freq - avg_pitch) < 10:  # Adjust the threshold as needed
            note = note_name
            break
    average_notes_every_132_frames.append(note)

# Display the average frequency and note for every 132 frames
print("Average Frequency and Notes for Every 132 Frames:")
for idx, (avg_pitch, avg_note) in enumerate(zip(average_pitch_every_132_frames, average_notes_every_132_frames)):
    print(f"Frames {idx * frame_interval + 1}-{(idx + 1) * frame_interval}: {avg_pitch:.2f} Hz, {avg_note}")
```

Implementation image 3.

```python
In [57]: audio_file = 'test.mp3'
         audio, sr = librosa.load(audio_file)

         # Compute the Short-Time Fourier Transform (STFT)
         stft = librosa.stft(audio)

         # Calculate the pitch for each frame
         frequencies, magnitudes = librosa.piptrack(S=stft)
         pitch_per_frame = np.nanargmax(magnitudes, axis=0)
         pitch_per_frame = [frequencies[i, t] for t, i in enumerate(pitch_per_frame)]

         # Define a mapping from frequencies to notes
         # Adjust these values based on the specific octave range in your audio
         note_mapping = {
             261.63: 'C4',
             277.18: 'C#4',  # Added C#
             293.66: 'D4',
             311.13: 'D#4',  # Added D#
             329.63: 'E4',
             349.23: 'F4',   # Added F
             369.99: 'F#4',  # Added F#
             392.00: 'G4',
             415.30: 'G#4',  # Added G#
             440.00: 'A4',
             466.16: 'A#4',  # Added A#
             493.88: 'B4',
             # Add more frequencies and notes as needed
         }
```

Implementation image 4.

```
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
```

Output 1

8 LIMITATIONS AND FUTURE WORK

8.1 Limitations

1. **Complex Music and Variability**: Current note detection systems may struggle with highly complex and fast-paced music compositions or those involving multiple instruments and overlapping notes. Developing algorithms capable of handling such scenarios remains a challenge.

2. **Instrument and Genre Specificity:** Many note detection algorithms are trained on specific instruments or genres, limiting their adaptability to a broader musical context. Future research should aim for more generalizable models.

3. **Real-time Processing:** Real-time applications, like live music analysis or interactive music software, often require low-latency processing, which can be a limitation in terms of computational resources and algorithm efficiency.

4. **Better accurate model:** Timbral variations in instruments and voices can impact note detection accuracy. Addressing these variations, particularly in polyphonic settings, remains an ongoing challenge.

8.2 Future Work

1. **Improved Accuracy**: Future research can focus on refining algorithms to enhance accuracy in complex musical scenarios. Leveraging deep learning techniques, such as neural networks, can lead to more robust note detection systems.

2. **Generalization:** Developing models that can recognize musical notes across various instruments and genres is a promising area of study. Transfer learning and domain adaptation techniques can be explored to achieve this.

3. **Real-time Applications:** Advancements in hardware and algorithm optimization can enable real-time note detection in live music performance and interactive applications.

4. **Expressive Analysis:** Going beyond basic note detection, future research can explore the analysis of musical expression and dynamics, providing a more nuanced understanding of music.

5. **Music Education:** Implementing note detection technology in music education tools can revolutionize how students learn and practice music.

6. **Interdisciplinary Integration:** Integration with fields like cognitive science, psychology, and neuroscience can lead to a deeper understanding of how humans perceive and create music.

7. **Data-Driven Music Research:** Note detection, when applied at scale, can facilitate large-scale musicological research, enabling data-driven insights into music history, trends, and cultural influences.

.

9 CONCLUSION

In our pursuit of understanding musical frequency note detection, this research has illuminated the vast potential for innovative applications and its profound impact on the fields of music and technology. The importance of accurate note detection across various domains, including music education, transcription, and audio processing, cannot be overstated. Precise note identification offers valuable tools for musicians and educators, enhancing the teaching and learning of music and enabling musicians to refine their performance and compositions. Additionally, our exploration has revealed how advancements in technology, such as digital signal processing and machine learning, have made automated note detection more efficient and accessible, paving the way for the development of software tools and applications that assist musicians, both amateur and professional, in their creative processes.

Furthermore, our research underscores the critical need for rigorous investigation in this field, emphasizing the significance of large and diverse datasets, robust algorithms, and the continuous refinement of techniques to improve the accuracy and reliability of note detection systems. In conclusion, the pursuit of musical frequency note detection transcends the technical realm, merging creativity and education with technology. This harmonious convergence between the worlds of music and technology offers new dimensions for artistic expression and learning. As we continue to refine and innovate, a future where music becomes more accessible, comprehensible, and enriched for all emerges on the horizon. This research serves as a glimpse into the vast potential that lies ahead in the realm of musical note detection.

References

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

[2] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S. -Y. Chang and T. Sainath, "Deep Learning for Audio Signal Processing," in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, pp. 206-219, May 2019, doi: 10.1109/JSTSP.2019.2908700.

[3] 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.

[4] 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.

[5] Hendrik Purwins , Bo Li , Tuomas Virtanen , Jan Schluter ¨ , Shuo-Yiin Chang, and Tara Sainath, “Deep Learning for Audio Signal Processing” in 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.

[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”.