KLE Technological University

Creating Value, Leveraging Knowledge

Dr. M. S. Sheshgiri Campus, Belagavi

Department of Electronics and Communication Engineering

Digital Signal Processing
Course Project
on

Lyrically - Automatic Synchronization of Textual Lyrics to Acoustic Music Signals

By:

Name	**SRN**
Sanskruti Mirajkar	02FE22BEC084
Shraman Kanthi	02FE22BEC092
Swati Patil	02FE22BEC112
Darshan Modekar	02FE22BEC119

Semester: V, 2024-2025

Under the Guidance of

Prof. Vinayak Dalavi

KLE Technological University,
Dr. M. S. Sheshgiri College of Engineering and Technology
BELAGAVI-590 008
2023-2024

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

CERTIFICATE

This is to certify that project entitled **"Lyrically - Automatic Synchronization of Textual Lyrics to Acoustic Music Signals "** is a bonafide work carried out by the student team of **" Sanskruti (02FE22BEC084) , Shraman (02FE22BEC092),Swati (02FE22BEC112),Darshan (02FE21BEC119)"**. The project report has been approved as it satisfies the requirements with respect to the Digital Signal Processing work prescribed by the university curriculum for B.E. (VI Semester) in Department of Electronics and Communication Engineering of KLE Technological University Dr. M. S. Sheshgiri CET Belagavi campus for the academic year 2023-2024.

Prof. Vinayak Dalavi **Guide**	**Dr. Dattaprasad A. Torse** **Head of Department**	**Dr. S. F. Patil** **Principal**

External Viva:

Name of Examiners **Signature with date**

1.

2.

ACKNOWLEDGMENT

Acknowledgment

We extend our heartfelt gratitude to our guide, **Prof. Vinayak Dalavi**, whose invaluable guidance and unwavering support have been instrumental throughout the entirety of this project. His expertise in digital signal processing has provided us with deep insights and clarity, enabling us to overcome challenges and achieve our project goals.Prof. Vinayak Dalavi comprehensive knowledge of spectral analysis, signal synchronization, and advanced DSP techniques has significantly contributed to the successful development of our system. His suggestions on techniques like Fourier Transform for spectral analysis, energy-based onset detection, and interpolation methods for synchronization have been crucial in ensuring the accuracy and robustness of our solution. Moreover, his emphasis on the importance of preprocessing, such as normalization, and his guidance on effective visualization methods have not only enhanced the technical quality of our work but also improved the overall presentation of our results. Prof. Dalavi's mentorship has fostered a spirit of learning and innovation, encouraging us to explore advanced concepts and refine our approach at every stage. Throughout this journey, his dedication, constructive feedback, and unwavering encouragement have been a source of motivation, inspiring us to push the boundaries of our understanding of DSP principles. His contributions have been vital in navigating the complexities of the project and ensuring its success.

We sincerely thank **Prof. Vinayak Dalavi** for his invaluable guidance, support, and encouragement, which have been the cornerstone of our efforts and achievements in this project.

-The project team

Contents

List of Figures

List of Abbreviations

Abbreviation	**Description**
DSP	Digital Signal Processing
FFT	Fast Fourier Transform
GUI	Graphical User Interface
STFT	Short-Time Fourier Transform
HMM	Hidden Markov Method

ABSTRACT

This paper describes the design and implementation of a system that synchronizes textual lyrics with acoustic music signals using advanced digital signal processing (DSP) techniques. The proposed system utilizes beat detection, spectral analysis, and temporal mapping of word timestamps to musical beats. The system aims to enhance multimedia applications such as karaoke systems, lyric video generation, and assistive tools for the hearing impaired by providing synchronized visual cues. Through iterative development, we learned to integrate MATLAB-based DSP functions with user-friendly graphical interfaces, ensuring the solution was both functional and engaging. The results highlighted the efficiency and versatility of DSP in solving complex problems, paving the way for further exploration in audio signal processing and its intersection with artificial intelligence and real-time applications.

1 INTRODUCTION

In the era of digital media and entertainment, music remains a universal mode of expression and storytelling. Lyrics are a vital component of music, conveying emotions, narratives, and artistic intent. However, manually synchronizing textual lyrics with acoustic music signals can be tedious and error-prone, especially for applications such as karaoke systems, music streaming platforms, and assistive technologies. To address this challenge, this project introduces **Lyrically: Automatic Synchronization of Textual Lyrics to Acoustic Music Signals**, a system that employs digital signal processing (DSP) techniques to automate this synchronization process.

The proposed system utilizes the inherent rhythmic and spectral properties of music to align lyrics with their corresponding audio signals in real time. The methodology involves analyzing the audio waveform to detect beats, a critical component for temporal alignment. The detected beats serve as anchors for mapping individual words from the lyrics to specific points in the music. By employing spectral analysis, energy-based beat detection, and temporal interpolation, the system ensures accurate alignment, even for dynamic and rhythmically complex audio.

The real-time performance and user-friendly graphical interface of this system make it suitable for a wide range of applications, including the creation of lyric videos, enhancing karaoke experiences, and assisting individuals with hearing impairments by providing synchronized visual cues. The modular design of the system also allows for potential extensions, such as multilingual lyric support and integration with advanced machine learning techniques.

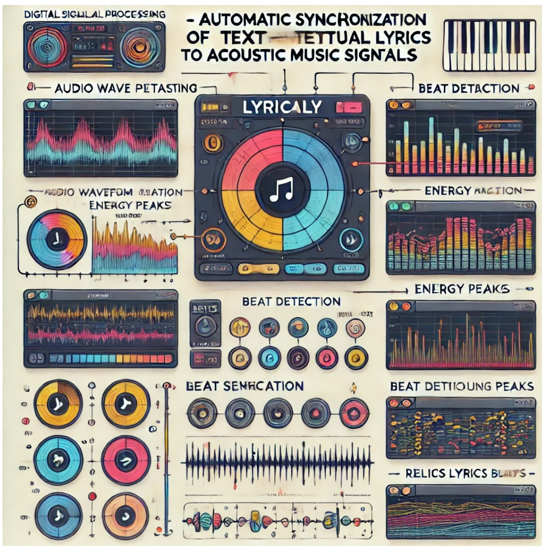

Figure 1.1: Overview of GUI

2 LITERATURE SURVEY

Sungkyun Chang (2017) proposed an unsupervised methodology to align song lyrics with their corresponding audio. The primary aim was to identify repetitive patterns in vowel acoustics and utilize them for effective alignment between lyrics and singing audio. The methodology relied on unsupervised learning techniques to analyze the acoustic properties of vowels, detecting repetitive patterns to establish the alignment without the need for labeled datasets or supervised training. However, the approach faced challenges in achieving high accuracy,

particularly in complex audio scenarios involving multiple instruments or overlapping vocals, which made isolating vowel patterns more difficult.
Yu-Ren Chien (2016) developed a robust model for aligning song lyrics with audio by leveraging acoustic and phonetic consistency in vowel sounds during singing. The research focused on improving alignment accuracy through acoustic-phonetic vowel likelihood modeling, which mapped the phonetic content of the lyrics to corresponding audio segments to ensure synchronization. Despite its advancements, the system struggled in handling polyphonic music, where multiple audio layers, such as instruments and background vocals, interfered with vowel detection. This limitation reduced the alignment accuracy in scenarios with complex musical arrangements.
Matthias Mauch (2011) enhanced the lyrics-to-audio alignment process by integrating chord information into a Hidden Markov Model (HMM) framework. By incorporating chord progression data, the methodology provided additional contextual features to refine the alignment process, particularly in complex musical compositions. Chords served as valuable features, improving the system's capability to match lyrics with their respective audio segments. However, the approach required highly accurate chord information, which can be difficult to obtain in real-world settings. Furthermore, unconventional chord progressions posed challenges for the HMM framework, leading to limitations in accurately modeling atypical harmonic structures.
Robert Macrae (2010) introduced "MuViSync," a prototype system designed for real-time synchronization of music audio and video. The objective was to ensure accurate playback synchronization by dynamically aligning audio and video streams. The system employed real-time feature detection algorithms to identify key features in both audio and video, which were then matched to achieve synchronization. Despite its real-time capabilities, the system faced processing limitations, particularly when handling high-resolution video and complex audio tracks simultaneously. Computational constraints further hindered its ability to achieve precise alignment in real-time environments.
Masataka Goto (2010) advanced the field of singing information processing by presenting techniques for singing voice modeling. The study aimed to improve the analysis and synthesis of audio by capturing subtle vocal characteristics, such as pitch, timbre, and dynamics. These models facilitated a deeper understanding of vocal nuances, enabling better audio synthesis and analysis. However, the approach encountered difficulties when dealing with diverse vocal styles, as the models struggled to generalize across different singers or genres. Additionally, noisy environments posed a significant challenge, reducing the system's accuracy in both analysis and synthesis tasks.

3 DATA SPECIFICATION

3.1 Textual Lyrics

The textual lyrics data consists of text files in formats such as .txt containing the complete lyrics of songs, repetitions. The lyrics are represented as sequences of words or phonemes that require pre-processing to remove irrelevant information for alignment purposes.

Sampled Lyrics

Song Name: Gayatri Stotra

OM BHUR BHUVAHASVAHA
TAT-SAVITUR VARENYAM
BHARGO DEVASYA DHIMAHI
DHIYO YONAH PRACHODAYAT

OM BHUR BHUVAHASVAHA
TAT-SAVITUR VARENYAM
BHARGO DEVASYA DHIMAHI
DHIYO YONAH PRACHODAYAT

3.2 Music Signals

The audio data used in our project is primarily in the WAV format, which is preferred for its uncompressed nature and ability to ensure high-quality audio analysis.The audio files have a standardized sampling rate of 44.1 kHz. The bit depth is typically 16-bit or 24-bit, depending on the audio source, with higher bit depths providing a better dynamic range to capture the nuances of singing vocals. The duration of audio files varies, usually ranging between 2 to 5 minutes, although longer tracks may be segmented into smaller portions for efficient processing.

4 METHODOLOGY

The methodology describes the design and implementation of a system that synchronizes lyrics with audio using Digital Signal Processing (DSP). The process includes hardware setup, data acquisition, and signal processing to analyze audio, detect beats, and map lyrics to timestamps.

4.1 Flow Chart

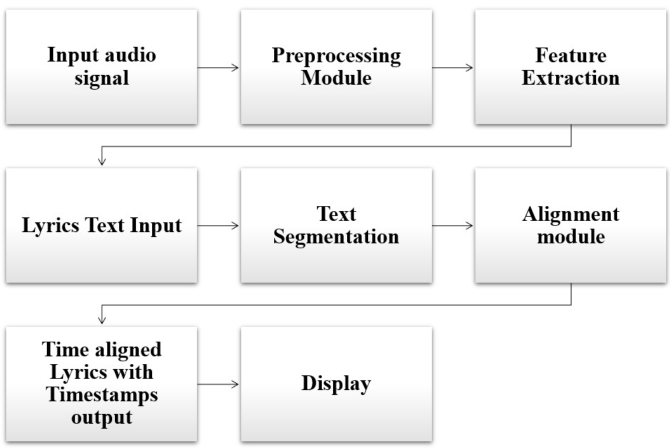

Figure 4.1: Flow chart

4.2 Data Acquisition

Audio Input

The user is prompted to load an audio file (.mp3 or .wav format) using the `uigetfile` function. The audio data is read using the `audioread` function. The audio is normalized to avoid amplitude inconsistencies, ensuring uniformity for further processing.

Lyrics Input

The user selects a text file containing the lyrics using the `uigetfile` function. The lyrics are split into individual words, and timestamps are generated based on beat detection.

4.3 Signal Processing

Beat Detection

Energy-Based Method

- The audio signal is divided into overlapping frames (20 ms with a 10 ms step).
- Energy for each frame is calculated by summing the squared amplitudes.
- Peaks in the energy signal are identified to determine beat times using MATLAB's `findpeaks` function.

Spectral Analysis

- The audio signal is divided into overlapping frames (20 ms with a 10 ms step).
- Energy for each frame is calculated by summing the squared amplitudes.
- Peaks in the energy signal are identified to determine beat times using MATLAB's `findpeaks` function.

Synchronization

- **Word Mapping:** Beat timestamps are interpolated to match the number of words in the lyrics. Each word is assigned a timestamp based on the detected beats.
- **Playback Synchronization:** During audio playback, lyrics are displayed in sync with the detected beats. A vertical marker dynamically updates on the waveform visualization to indicate the current timestamp.

5 GUI Implementation

The waveform visualization component plots the audio waveform on an axes object, with amplitude represented on the y-axis and time on the x-axis. Additionally, the spectral display showcases the frequency spectrum of the audio in a separate axes object, providing insight into the audio's harmonic content. For beat detection representation, detected beats are marked on an energy-time plot, using red circles to indicate peak positions, which facilitates the identification of rhythmic patterns within the audio. Finally, the lyrics display features a text box in the GUI that presents the synchronized lyrics word by word during playback, ensuring that viewers can follow along with the music seamlessly.

5.1 Control Buttons

The "Load Audio" function prompts the user to select an audio file and subsequently visualizes the waveform along with the spectral analysis of the audio. Similarly, the "Load Lyrics" feature allows the user to load a lyrics file, processing the text to map words to their corresponding timestamps. Once the audio and lyrics are loaded, the "Play" function initiates audio playback while synchronizing the lyrics and updating the visualizations in real-time. Conversely, the "Stop" function halts audio playback and resets the graphical user interface (GUI), allowing users to restart the process as needed.

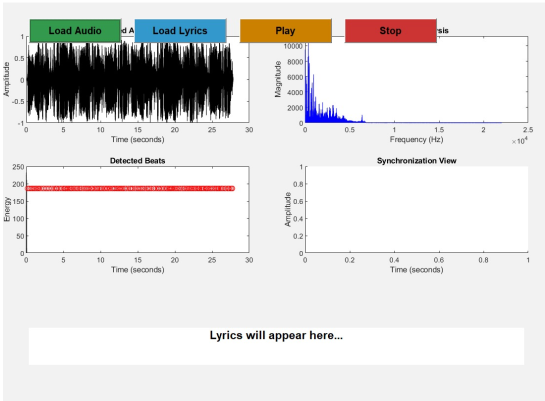

Figure 5.1: GUI Interface

6 RESULTS

6.1 Waveform and Spectrogram Visualization

The visualization of the audio waveform (amplitude vs. time) and the spectrogram (frequency vs. time) confirms the accurate processing of audio signals.

- Clear representation of beats and transitions in the waveform.
- Frequency components are effectively highlighted in the spectrogram for diverse audio samples.

Additionally, user controls such as "Load Audio," "Play," and "Stop" provide an interactive and seamless experience for analyzing audio data.

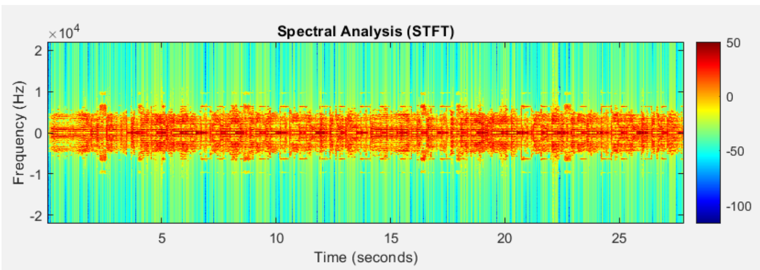

Figure 6.1: Spectrum Analysis

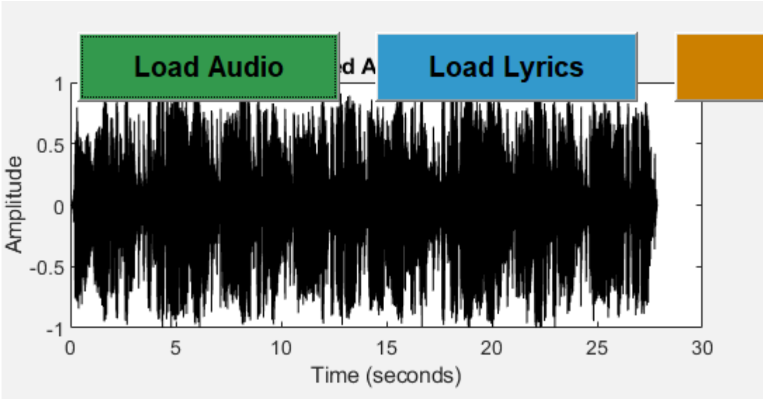

Figure 6.2: Normalized waveform

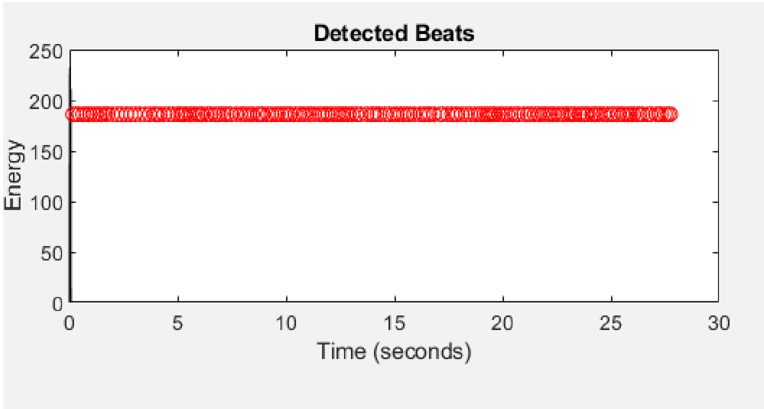

Figure 6.3: Detected Beats

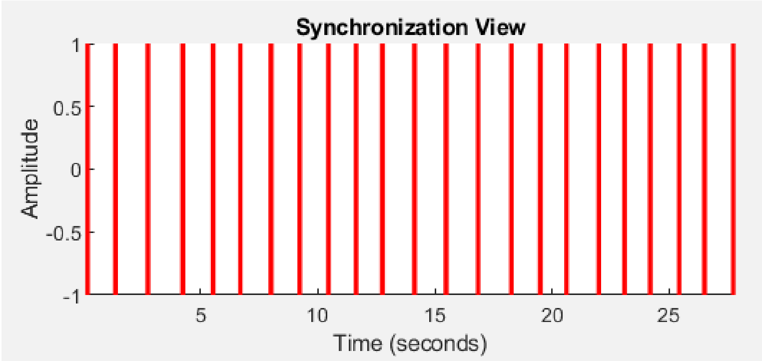

Figure 6.4: Synchronization View

Figure 6.5: Lyrics Syncronized with audio

7 APPLICATIONS

- **Music Education**: The system can be used as a learning tool for music students, allowing them to follow along with lyrics while listening to songs. This can help improve their understanding of timing and pronunciation in singing.
- **Karaoke Systems**: The synchronization technology can enhance karaoke experiences by displaying lyrics in real time with the music, helping users to sing along accurately and in time with the audio.
- **Lyric Video Creation**: Content creators can utilize the system to automatically generate lyric videos for songs, aligning the textual lyrics with the corresponding audio tracks for a visually appealing presentation.
- **Music Analysis**: Researchers and musicologists can use the system for analyzing the structure of songs, studying the relationship between lyrics and music, and exploring how different sections of a song are emphasized in the audio.
- **Accessibility Tools**: The project can aid individuals with hearing impairments by providing a visual representation of lyrics in sync with music, making it easier for them to engage with musical content.
- **Interactive Media**: In applications like video games or interactive storytelling, the synchronization of lyrics with music can enhance user engagement by providing a seamless experience of audio-visual integration.

8 ADVANTAGES

- It significantly improves the accessibility of music for individuals with hearing impairments by providing synchronized visual representations of lyrics, allowing them to engage more fully with musical content.
- The system serves as a valuable educational tool for music students, helping them understand timing, pronunciation, and lyrical structure while practicing their singing.
- It enhances karaoke experiences by ensuring that lyrics are displayed in real time with the music, making it easier for users to sing along accurately.
- For content creators, the project facilitates the automatic generation of engaging lyric videos, streamlining the production process while ensuring high-quality alignment between audio and text.
- Researchers and musicologists benefit from the system by utilizing it to analyze the relationship between lyrics and music, thereby gaining insights into song structures and their emotional impacts.

- Overall, this project not only enriches the music experience for end-users but also provides useful tools for educators, creators, and researchers alike.

9 FUTURE WORK

- Incorporating advanced machine learning algorithms to enhance the accuracy of beat detection and lyric synchronization, particularly in complex musical pieces with varying tempos and styles.
- Integrating natural language processing techniques to better understand and handle different lyrical structures, including those with complex rhymes and unconventional phrasing.
- Expanding the range of supported audio formats and improving the system's robustness against background noise, allowing for greater versatility in real-world applications.
- Developing a mobile application version to make the technology more accessible to a wider audience, enabling users to sync lyrics with their favorite songs on-the-go.
- Exploring collaborative features that allow multiple users to engage with the system simultaneously, opening up new avenues for interactive learning and performance.

CONCLUSION

Throughout the project, we utilized various DSP methods to process, analyze, and synchronize audio signals with textual lyrics effectively. Techniques such as **Fourier Transform** for spectral analysis, **energy-based onset detection** for beat and rhythm identification, and **interpolation methods** for mapping lyrics to detected beats played crucial roles in achieving accurate results. Preprocessing steps like normalization ensured robust handling of diverse audio inputs, while visualization tools offered intuitive representations of waveform and spectral features.

REFERENCES

1. *Kan, Min-Yen, et al. "LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals." IEEE Transactions on Audio, Speech, and Language Processing, 16(2), 2008, pp. 338-349.*
2. *Kan, M. Y., Wang, Y., Iskandar, D., Nwe, T. L., Shenoy, A. (2008). LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals. IEEE Transactions on Audio, Speech, and Language Processing, 16(2), 338-349.*
3. *Kan, Min-Yen, Ye Wang, Denny Iskandar, Tin Lay Nwe, and Arun Shenoy. "LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals." IEEE Transactions on Audio, Speech, and Language Processing, 16(2), 2008, pp. 338-349.*
4. *Kan, M.Y., Wang, Y., Iskandar, D., Nwe, T.L., and Shenoy, A. (2008). LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals. IEEE Transactions on Audio, Speech, and Language Processing, 16(2), pp. 338-349.*
5. *Kan, M.Y., Wang, Y., Iskandar, D., Nwe, T.L., Shenoy, A. LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals. IEEE Transactions on Audio, Speech, and Language Processing, 16(2), 2008, pp. 338-349.*