A
Mini-Project Report on

MUSIC STREAMING APP WITH MULTILINGUAL FEATURES

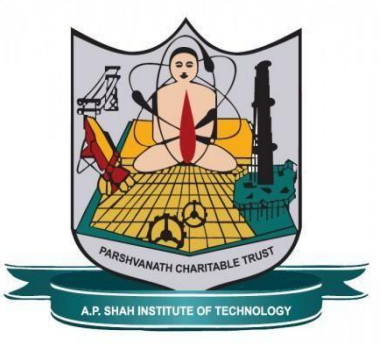
Submitted in fulfillment of the requirements for the degree of
BACHELOR OF ENGINEERING
IN
Computer Science & Engineering
(Artificial Intelligence & Machine Learning)

by

Swayam Bhoir (23106025)
Rushi Choraghe (23106033)
Arif Choudhary (23106054)
Nishant Dakua (23106014)

Under the guidance of

Prof. Sayali Badhan



Department of Computer Science & Engineering
(Artificial Intelligence & Machine Learning)
A. P. Shah Institute of Technology
G. B. Road, Kasarvadavali, Thane (W)-400615
University Of Mumbai
2024-2025

A. P. SHAH INSTITUTE OF TECHNOLOGY

CERTIFICATE

This is to certify that the project entitled "**Music Streaming Web App with Multilingual Features"** is a Bonafide work of Rushi Choraghe (23106033), Swayam Bhoir (23106025), Arif Choudhary(23106054), Nishant Dakua (23106014) submitted to the University of Mumbai in fulfillment of the requirement for the award of **Bachelor of Engineering** in **Computer Science & Engineering (Artificial Intelligence & Machine Learning).**

Prof. Sayali Badhan	Dr. Jaya Gupta
Mini Project Guide	Head of Department

A. P. SHAH INSTITUTE OF TECHNOLOGY

Project Report Approval

This Mini project report entitled "**Music Streaming Web App with Multilingual Features"** by **Swayam Bhoir , Arif Choudhary, Nishant Dakua and Rushi Choraghe** is approved for the degree of ***Bachelor of Engineering*** in ***Computer Science &Engineering*** **(AI & ML),** ***2024-25***.

External Examiner: _____

Internal Examiner: _____

Place: APSIT, Thane
Date:

Declaration

We declare that this written submission represents my ideas in my own words and where others' ideas or words have been included, I have adequately cited and referenced the original sources. I also declare that I have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission. I understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission hasnot been taken when needed.

Swayam Bhoir	Rushi Choraghe	Arif Choudhary	Nishant Dakua
(23106025)	(23106033)	(23106054)	(23106014)

ABSTRACT

The rise of global music streaming platform has enabled user to explore diverse musical genres across different languages. However, language barrier often hinders full appreciation of song lyrics and meaning. This project aims to develop a web-based music streaming application with an innovative feature: real-time Hindi subtitles for English songs. The application will provide seamless music streaming with a user-friendly interface, allowing users to access a vast collection of songs from various languages while displaying synchronized English subtitles. These subtitles will be generated through a combination of AI-based translation, crowdsourced lyrics, and official lyric databases, through an Natural Language Processing (NLP) model ensuring accuracy and context preservation. The platform will be developed using modern web technologies for high performance and cross-platform compatibility. The backend will manage audio processing, subtitle synchronization, user preferences, and content recommendations, while the frontend will offer an intuitive experience for users. By bridging linguistic gaps, this application aims to enhance global music accessibility, enabling users to enjoy and understand songs beyond their native language.

Keywords: AI-based Translation, Crowdsourced Lyrics, NLP Model, Transformers

Index

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

The rapid growth of digital technology has transformed the music industry, making music streaming platforms an integral part of how people access and enjoy music globally. These platforms have broken geographical barriers, allowing users to explore and experience songs from different cultures and languages. However, despite this accessibility, a significant challenge remains—language barriers often prevent users from fully understanding and appreciating music in foreign languages. Recognizing this gap, our web-based music streaming application introduces an innovative feature: the automatic generation of English subtitles for songs in various languages.

This platform is designed to provide a seamless and immersive music streaming experience while addressing the linguistic challenges that many listeners face. By integrating real-time English subtitles, users can understand the meaning of lyrics in their preferred language, enabling them to connect with songs beyond just their melodies. This feature is particularly beneficial for individuals who enjoy international music, language learners, and those who wish to gain deeper insights into the cultural and poetic elements embedded in foreign-language songs.

The core functionality of this music streaming platform revolves around advanced natural language processing (NLP) and machine learning algorithms that facilitate real-time lyric translation and synchronization. Using AI-driven techniques, the application extracts lyrics from songs, translates them accurately into English, and aligns them with the corresponding audio. The subtitle generation process ensures that translations maintain the contextual meaning and artistic essence of the original lyrics. This approach not only enhances user engagement but also provides a richer and more meaningful listening experience.

From a technical perspective, the platform incorporates a well-structured backend architecture that supports high-quality music streaming, real-time subtitle display, and seamless user interaction. The integration of cloud-based services ensures smooth performance and scalability, allowing the application to handle a vast database of multilingual songs efficiently. Additionally, an intuitive and user-friendly interface enhances accessibility, enabling users to easily search for songs, customize subtitle settings, and enjoy a personalized.

Music transcends linguistic barriers, evoking emotions and connecting cultures through its

universal language of melody and rhythm. However, the nuanced meanings and cultural context embedded within song lyrics often remain inaccessible to those who do not understand the original language. This report details the development and evaluation of a novel music lyrics translation model, aimed at bridging this linguistic divide and fostering a deeper appreciation for global musical artistry.

The increasing globalization of music consumption necessitates accurate and culturally sensitive translation of song lyrics. Existing translation methods, often relying on literal word-for-word conversion, frequently fail to capture the poetic essence, emotional undertones, and idiomatic expressions inherent in lyrical content. This can lead to misinterpretations and a diminished understanding of the artist's intended message. Our research addresses this challenge by exploring advanced natural language processing (NLP) techniques specifically tailored for the complexities of music lyrics.

This report outlines the methodology employed in building our translation model, encompassing data collection, preprocessing, model architecture selection, training, and evaluation. We delve into the unique characteristics of music lyrics, such as their rhythmic structure, rhyming schemes, and figurative language, and how these factors influence the design of our model. Furthermore, we explore the challenges associated with preserving the artistic intent and cultural nuances during the translation process.

Our approach leverages state-of-the-art sequence-to-sequence models, incorporating attention mechanisms and contextual embeddings to capture the intricate relationships between words and phrases within the lyrical context. We also investigate the integration of external knowledge sources, such as cultural databases and musical genre-specific lexicons, to enhance the accuracy and cultural appropriateness of the translations.

The evaluation of our model's performance is conducted through a combination of automatic metrics and human evaluation. We compare our model's output against existing translation systems and human-generated translations, focusing on aspects such as semantic accuracy, fluency, preservation of poetic devices, and cultural sensitivity. The findings of this evaluation provide valuable insights into the strengths and limitations of our approach, paving the way for future improvements and advancements in the field of music lyrics translation. Ultimately, this research contributes to a more inclusive and accessible global music landscape, enabling a wider audience to connect with the rich tapestry of musical storytelling from around the world.

CHAPTER 2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 HISTORY

The evolution of music streaming began with the disruptive arrival of Napster in 1999, which pioneered pe er-to-peer file sharing. Though plagued by legal battles, it fundamentally altered music consumption. The early 2000s saw the rise of digital downloads, with Apple's iTunes Store establishing a model for legal online music sales. Platforms like Pandora introduced personalized internet radio, shaping how listeners discovered new music.

Spotify's 2008 launch marked a shift towards on-demand streaming, offering vast libraries accessible through subscriptions. Apple Music followed, further solidifying streaming's dominance.

Today, services like YouTube Music and Tidal cater to diverse needs, from video integration to high-fidelity audio. Streaming has revolutionized music distribution, impacting artists, labels, and how we experience music globally.

Existing Websites and Their Features

The online music streaming landscape is dominated by a few key players, each offering a range of features to cater to diverse listening preferences. Here's a breakdown of common features and some of the major platforms:

- Vast Music Libraries
- On-Demand Streaming
- Personalized Playlists
- Offline Listening
- Search and Discovery
- High Quality Audio
- Lyrics Display

Challenges and Opportunities:

A major point of contention is the low per-stream payout to artists. Many musicians feel they are not adequately compensated for their work, leading to ongoing debates about fair royalty distribution. Balancing profitability for the platforms with fair compensation for artists is a constant struggle.

2.2 LITERATURE REVIEW

- **Jurafsky, D., & Martin, J. H. Speech and Language Processing, Stanford University (2021):** This paper Describes machine translation (MT), the use of computers to translate from one language to another. Of course translation, in its full generality, such as the translation of literature, or poetry, is a difficult, fascinating, and intensely human endeavor, as rich as any other area of human creativity. Machine translation in its present form therefore focuses on a number of very practical tasks. Perhaps the most common current use of machine translation is for information access. We might want to translate some instructions on the web, information access perhaps the recipe for a favorite dish, or the steps for putting together some furniture. Or we might want to read an article in a newspaper, or get information from an online resource like Wikipedia or a government webpage in some other language. MT for information access is probably one of the most common uses of NLP technology. [1]

- **Fujihara, H., Goto, M, Lyric Synchronization Using Music Audio Alignment and Text Processing. IEEE Transactions on Audio, Speech, and Language Processing (2021):** This paper presents a system that automatically synchronizes lyrics with music audio, using alignment techniques and text processing for accurate lyric display. Automatically syncing lyrics is difficult, particularly with songs that have fast lyrics or varying tempos. It is hard to accurately determine the correct timing of each word or phrase in relation to the audio. In the last decade, there has been considerable interest in digital music services that display the lyrics of songs that are synchronized with their audio. As a recent example, Sound Hound. Although some well-known hit songs already have lyric time stamps within existing karaoke databases, more than a hundred million numbers, including new, unpopular songs, cover songs, and live recordings in YouTube. [2]

- **Vaswani, A., Shazeer, N., Parmar. Attention Is All You Need. Advances in Neural Information Processing Systems (2017):** This book describes the Recurrent neural networks, especially LSTMs and GRUs, are state-of-the-art for sequence modeling, with ongoing efforts to improve them. Pushing the boundaries of recurrent language models and encoder-decoder architectures. Efficient inference and visualization of models is tedious to maintain. The goal of reducing sequential computation also forms the foundation of the Extended Neural GPU. Self-attention, sometimes called intra-attention is an attention mechanism relating different

positions of a single sequence in order to compute a representation of the sequence. Self-attention has been used successfully in a variety of tasks including reading comprehension, abstractive summarization, textual entailment and learning task-independent sentence representations. [3]

- **Peeters, G. A Large Set of Audio Features for Sound Description: Audio presentation for Music Information Retrieval (2020):** This paper proposes an adaptive acoustic anomaly detection system using optimized dimension reduction and clustering, achieving high accuracy and efficiency in industrial machine monitoring. The core challenge is to identify anomalous machine conditions before breakdowns occur. Finding the best parameters for clustering and dimension reduction across diverse environments is difficult. In modern industry, maintaining continuous machine operations is important for improving production efficiency and reducing costs. Therefore, the smart technology of acoustic monitoring to detect anomalous machine conditions earlier before breakdowns works as part of predictive maintenance and is applied not only in industry fault detection but also in safety monitoring and surveillance systems. [4]

- **Marmik Pandya, NLP Based Poetry Analysis and Generation, Norteastern University (2016):** This paper describes NLP techniques tailored to the specific characteristics of poetry and literary text. Computational analysis of poetry is still an active research problem area. Transferring these techniques to song lyrics requires adaptation. Natural Language Generation is a popular research field in NLP and there are many approaches right from using Markov chains and context free grammar to using recurrent volume networks and other deep learning techniques proposed to make computer generate poetry. This study tries analyze a few computational models being suggested for analysing poetry and also discuss a few of the poetry generation approaches [5]

- **Anders Søgaard, Ivan Vuli´c, Sebastian Ruder, Manaal Faruqui, Cross-Lingual Word Embeddings, University of the Basque Country, (2019):** This book talks about Learning representation of words that capture semantic learning across languages. These are improving MT and other cross lingual tasks. Capturing the nuances of musical and emotional context is difficult. Limited resources for some languages pairs is relatively well resourced but still presents challenges. The interest in cross-lingual word embeddings has grown in recent years. This is partly because of their success in cross-lingual transfer, where NLP tools trained in a resource-rich language such as English are transferred to another language with smaller or no

annotated data. For instance, given training data for a text-classification task in English, a model using CLWE can classify foreign language documents. Beyond language pairs, CLWE allows us to represent words of several languages in a common space, and thus pave the way to build multilingual NLP tools that use the same model to process text in different languages. [6]

Key Considerations for App:

- **Hybrid Approach:** Combination of automatic translation (using a good NMT system like Google Translate or Deep Learning as a base) with post-editing and refinement, potentially crowdsourced or by experts.
- **Language-Specific Fine-tuning:** If possible, fine-tune your chosen MT system on a dataset of song lyrics (even a small, curated one would help).
- **Handling Poetic Devices:** Developing strategies for handling rhyme, rhythm, and meter. This might involve rule-based systems or templates.
- **User Interface:** This project provides a way for users to suggest corrections or improvements to the translated lyrics.

CHAPTER 3

PROBLEM STATEMENT

3. PROBLEM STATEMENT

Music is a universal language, yet language barriers often prevent listeners from fully understanding and appreciating songs in different languages. While existing music streaming platforms provide lyrics for some songs, few offer real-time Hindi subtitles for English tracks, making it difficult for global audiences to connect with the meaning behind the music. As a result, listeners who enjoy foreign music often have to rely on third-party translation tools, disrupting the immersive experience. Additionally, current music streaming platforms lack advanced personalization and language-based recommendations that cater to users who enjoy multilingual music. Many platforms focus primarily on mainstream content, making it harder for users to discover international artists and culturally diverse music. Moreover, seamless user interaction and accessibility features, such as lyric synchronization, language toggling, and intelligent recommendations, are often missing from existing solutions. Our project aims to develop a web-based music streaming application that not only provides high-quality music playback but also integrates real-time Hindi subtitles for English songs. This will enable users to fully experience and understand the meaning of the lyrics while listening. Additionally, the platform will feature personalized recommendations, intelligent language-based playlists, and an interactive interface to enhance user engagement. By addressing these limitations, our application will serve as a bridge between cultures, making music from different parts of the world more accessible, inclusive, and enjoyable. Whether a user is a language learner, a fan of foreign music, or simply someone looking to explore new sounds, our platform will offer a unique and enriching experience beyond standard music streaming services.

\

CHAPTER 4

EXPERIMENTAL SETUP

4. EXPERIMENTAL SETUP

4.1 Software Setup

- **NodeJS:** The runtime is required to handle the requests and the managing the backend from the server. Node.js is a JavaScript runtime environment built on Chrome's V8 JavaScript engine. Essentially, it allows you to run JavaScript code outside of a web browser. This was a significant shift as JavaScript was traditionally limited to front-end web development.

- **Firebase:** Firebase is used for user authentication and databases. Firebase is a comprehensive Backend-as-a-Service (BaaS) platform provided by Google. It's designed to help developers build, manage, and grow their mobile and web applications without needing to manage the underlying server infrastructure. Think of it as a suite of ready-to-use tools and services that handle many common backend tasks, allowing you to focus on building the frontend and user experience of your app.

- **React:** To build different frontend Components. React is a declarative, efficient, and flexible JavaScript library for building user interfaces (UIs) or UI components. It's maintained by Meta (formerly Facebook) and a large community of individual developers and companies.

- **Frontend:** Developed using React.js to create a responsive and dynamic user interface. Usually creating frontend only using HTML and CSS can be a tedious task, for that reason React is used. Frontend is composed of multiple react componants. Using such frameworks helps in creating a single page application. Which means, all the assets are loaded at once, to avoid delay later.

- **Backend:** Backend development using Node.js and Firebase is a popular and efficient approach for building scalable and real-time applications. The combination leverages the strengths of both technologies.

- **Database:** Firebase Firestore for storing user preferences, playlists, and translated lyrics. The lyrics are stored in the firestore database to ensure reduction of delays.

- **Music APIs:** Integrated with custom API with firebase firestore. Music database is stored in the firebase, as no there does not exist any free API to fetch music and its metadata.

4.2 Hardware Requirements

- **Processor:** While an i5 processor (or higher) is generally recommended for a decent computing experience, it's not strictly necessary for training all NLP models, especially smaller ones or for initial experimentation. However, there are several compelling reasons why a more powerful processor like an i5 or higher becomes highly beneficial and often essential for more demanding NLP tasks:

- **RAM:** While 4 GB of RAM is severely limiting for most NLP training tasks, 8 GB of RAM represents a more reasonable starting point for basic experimentation and working with smaller to medium-sized datasets and less complex models.

- **GPU**: NVIDIA GPU (Optional but recommended for AI model training)

- **Internet:** Stable speed connection for API calls and cloud interactions

The experimental setup ensures efficient data processing, seamless integration, and high-performance execution of the music streaming and translation functionalities.

CHAPTER 5

PROPOSED SYSTEM & IMPLEMENTATION

5. PROPOSED SYSTEM & IMPLEMENTATION

5.1 Block diagram of proposed system

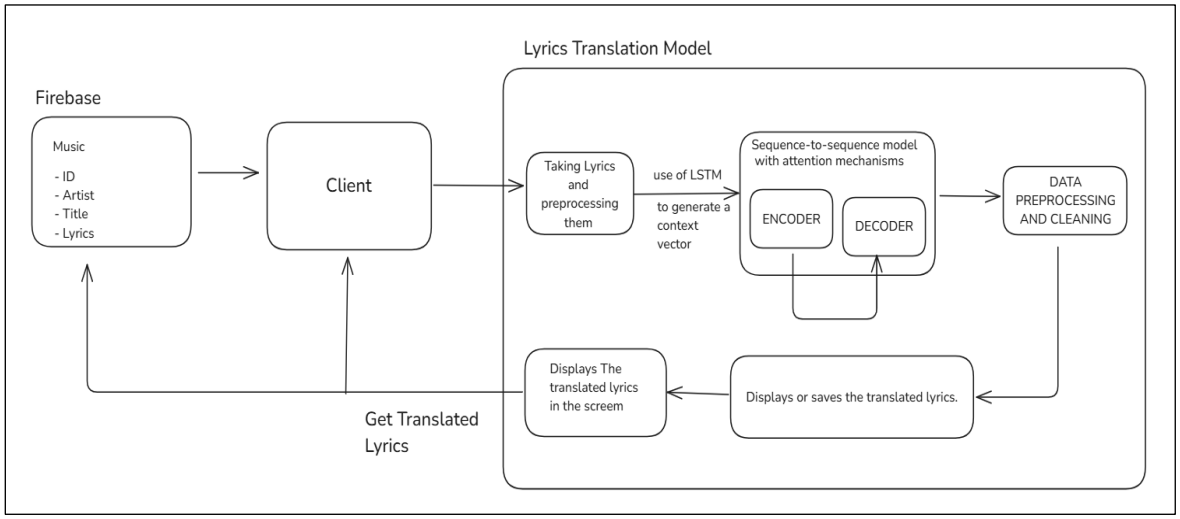


Figure 5.1 Lyrics Translation Model

5.2 Description of block diagram

Firebase: Firebase is a comprehensive Backend-as-a-Service (BaaS) platform provided by Google. It's designed to help developers build, manage, and grow their mobile and web applications without needing to manage the underlying server infrastructure. Think of it as a suite of ready-to-use tools and services that handle many common backend tasks, allowing you to focus on building the frontend and user experience of your app.

Lyrics Translation Model:

- **Sequence-to-Sequence Architecture:** The core of the model is a sequence-to-sequence (seq2seq) architecture. This is a common structure for tasks like machine translation, where an input sequence (lyrics in one language) is transformed into an output sequence (lyrics in another language).

- **LSTM (Long Short-Term Memory):** The model employs LSTM units. LSTMs are a type of recurrent neural network (RNN) designed to handle long-range dependencies in

often rely on relationships between words that are far apart in the sequence.

- **Encoder:** The encoder processes the input lyrics and transforms them into a fixed-length context vector. This vector is a compressed representation of the input lyrics, capturing their essential information. The encoder's role is to read and process the entire input sequence (the sentence in the source language) and encode it into a fixed-length vector (also known as the context vector, thought vector, or sentence embedding).

- **Decoder:** The decoder takes the context vector generated by the encoder and uses it to generate the translated lyrics. It does this word by word, using the attention mechanisms to focus on relevant parts of the input.

- **Data Preprocessing and Cleaning:** Before being fed into the model, the lyrics data undergoes preprocessing and cleaning. It is important to improve performance of the model. Data preprocessing is a crucial and often the most time-consuming step in any data science or machine learning project. It involves transforming raw data into a clean, organized, and suitable format that can be effectively analyzed and used to train machine learning models. Raw data is often messy, incomplete, inconsistent, and contains noise, which can lead to inaccurate results and poorly performing models if not handled properly.

- **Output:** The model can either display the translated lyrics directly on the screen or save them to the cloud database, like in this case Firestore.

CHAPTER 6

CONCLUSION

6. CONCLUSION

The rapid growth of digital technology has transformed the music industry, making music streaming platforms an integral part of how people access and enjoy music globally. These platforms have broken geographical barriers, allowing users to explore and experience songs from different cultures and languages. However, despite this accessibility, a significant challenge remains—language barriers often prevent users from fully understanding and appreciating music in foreign languages. Recognizing this gap, our web-based music streaming application introduces an innovative feature: the automatic generation of English subtitles for songs in various languages. This feature not only enhances the user experience by providing deeper engagement with diverse musical content but also fosters a greater appreciation for cultural nuances embedded within the lyrics. By bridging the linguistic divide, our application opens up a world of musical understanding, allowing listeners to connect with the emotional and narrative depth of songs regardless of their original language.

REFERENCES

Research Papers & Articles

[1] Jurafsky, D., & Martin, J. H. (2021). Speech and Language Processing. Stanford University.

[2] Fujihara, H., Goto, M., Ogata, J., & Okuno, H. G. (2021). Lyric Synchronization Using Music Audio Alignment and Text Processing.

[3] Vaswani, A., Shazeer, N., Parmar, N., et al. (2020). Attention Is All You Need. Advances in Neural Information Processing Systems.

[4] Peeters, G. (2017), A Large Set of Audio Features for Sound Description: Audio presentation for Music Information Retrieval

[5] Marmik Pandya (2016), NLP Based Poetry Analysis and Generation, Northeastern University

[6] Anders Søgaard, Ivan Vuli´c, Sebastian Ruder, Manaal Faruqui (2019), Cross-Lingual Word Embeddings, University of the Basque Country.

Links

- Spotify Developer API https://developer.spotify.com
- Musixmatch API https://developer.musixmatch.com
- YouTube Captions https://support.google.com/youtube/answer/6373554