AI-BASED AUTOMATIC MUSIC COMPOSITION VOCAL SYNTHESIS

Y.Nagamalleswararao [1],K.Pavani[2],K.Akanksha Sai[3]
#1 Assistant Professor in the Department Of MCA, SRK Institute Of Technology,Vijayawada.

#2 Assistant Professor & Head of Department of MCA, SRK Institute of Technology, Vijayawada.

#3 Student in the Department of MCA, SRK Institute of Technology, Vijayawada

.

Abstract: The rapid advancement of Artificial Intelligence and Natural Language Processing has enabled innovative applications in music generation and audio synthesis. This project proposes an AI-Based Automatic Music Composition Vocal Synthesis System Using Text, which transforms user-provided text into complete musical compositions with synthesized vocals, melodies, harmonies, bass lines, and rhythmic accompaniment automatically.

The system processes textual input through Natural Language Processing techniques to analyze lyrics and musical characteristics. Using AI-driven melody generation and audio synthesis, it creates songs in multiple languages, including English, Hindi, Telugu, Tamil, French, and Spanish. Users can select male or female voices, while audio normalization techniques ensure high-quality sound output.

The proposed system follows a three-tier architecture consisting of a React.js frontend, a Python FastAPI backend, and an AI-powered music generation module. The frontend provides an interactive user interface with real-time visualization, while the backend manages processing, audio generation, and file handling. The AI module generates melodies, chord progressions, bass patterns, and vocal tracks, which are combined into a final musical composition. The system can generate complete songs within seconds, significantly reducing manual effort in music production. This approach makes AI-powered music creation accessible to students, content creators, independent artists, and educational institutions, providing a scalable and user-friendly solution for automatic music composition.

Index terms *-Artificial Intelligence, Natural Language Processing, Text-to-Music, Automatic Music Composition, Vocal Synthesis, Melody Generation, Audio Processing, Deep Learning, Multi-language Music Generation, FastAPI, React.js..*

1. INTRODUCTION

Music is one of the most powerful forms of human expression and plays an important role in entertainment, education, communication, and cultural preservation. Traditionally, creating music requires knowledge of musical theory, composition techniques, vocal recording, and audio production tools. This process can be time-consuming and often requires professional expertise. With the rapid advancement of Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP), it has become possible to automate many aspects of music creation.

The project, "Automatic Music Composition and Vocal Synthesis Using Text," aims to develop an intelligent system that can generate complete musical compositions from simple text input provided by the

user. The system automatically converts textual content into songs by generating vocals, melodies, harmonies, bass lines, and rhythmic patterns without requiring any musical knowledge from the user.

The The proposed system utilizes Natural Language Processing techniques to analyze the input text and identify its structure, language, and musical characteristics. Based on this analysis, the system creates suitable melodies and instrumental arrangements. Advanced Text-to-Speech technology is used to synthesize realistic human-like vocals in multiple languages, while audio processing algorithms combine vocals and background music into a final song composition.

The application follows a modern full-stack architecture consisting of a React.js frontend, a Python FastAPI backend, and an AI-powered music generation engine. The user interacts through a simple web interface where text can be entered, language and voice preferences can be selected, and the generated song can be played or downloaded instantly.

This project demonstrates how AI can be applied in creative domains to transform textual ideas into complete musical experiences, making automated music composition more efficient, scalable, and user-friendly.

2. LITERATURE SURVEY

3.1 Automatic Music Composition using Artificial Intelligence (2019)

In 2019, a research study published in IEEE explored the use of Artificial Intelligence for automatic music composition. The researchers developed a system capable of generating melodies using machine learning algorithms trained on musical datasets. The system analyzed musical patterns, note sequences, and rhythm structures to create new compositions automatically.The experimental results demonstrated that AI-based music generation systems can assist users in creating original musical content while reducing the time and effort required for manual composition.

3.2 Deep Learning Based Music Generation using Neural Networks (2020)

In 2020, researchers proposed a deep learning-based music generation system using Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures. The objective of the study was to generate realistic melodies and musical sequences by learning patterns from existing musical compositions.The system was trained using publicly available music datasets and demonstrated the ability to generate coherent melodies and rhythmic structures. The study concluded that deep learning techniques significantly improve the quality and creativity of automatically generated music.

3.3 Text-to-Speech Based Vocal Synthesis Systems (2020)

A Another research study focused on advanced Text-to-Speech technologies for vocal synthesis. The proposed system utilized neural speech synthesis techniques to convert textual input into natural-sounding human speech. Multiple languages and voice styles were supported to improve user experience.The research findings indicated that modern Text-to-Speech systems can generate highly realistic vocals and play a significant role in music production and audio content generation.

3.4 Transformer-Based Music Generation Models (2021)

In 2021, researchers introduced Transformer-based architectures for automatic music composition. The system used attention mechanisms to learn long-range musical dependencies and generate complex melodies and harmonies.The model was evaluated using large-scale music datasets and produced high-quality compositions with improved consistency. The study demonstrated that Transformer-based models are highly effective for AI-driven music generation applications.

3.5 AI-Powered Multilingual Music Generation Framework (2022)

In 2022, a multilingual music generation framework was proposed for creating songs in

different languages. The system integrated Natural Language Processing, Text-to-Speech synthesis, and AI-based melody generation techniques..

3. METHODOLOGY

i) Proposed Work:

The proposed AI-Based Automatic Music Composition and Vocal Synthesis System automatically converts user-provided text into complete musical compositions. The system uses Artificial Intelligence, Natural Language Processing, and Audio Synthesis techniques to generate vocals, melodies, harmonies, and rhythmic patterns. It supports multiple languages and voice options for personalized song creation. The system consists of a frontend interface, backend services, and an AI processing module. The AI engine analyzes text, generates music, synthesizes vocals, and combines all audio components into a final song. The proposed system reduces manual effort, simplifies music production, and makes song creation accessible to users without musical expertise.

ii) System Architecture:

The architectural design of the Automatic Music Composition and Vocal Synthesis Using Text System explains how different parts of the system are organized and how they work together to generate complete musical compositions automatically. The system is designed in a structured and modular manner so that each component performs a specific function efficiently.The architecture mainly consists of the frontend interface, backend server, AI music generation module, audio processing module, and database system. These components communicate with each other using REST APIs and JSON data exchange. The design helps maintain scalability, reliability, efficiency, and high-quality music generation throughout the application.

Fig1 proposed architecture

iii) Modules:

The proposed Automatic Music Composition and Vocal Synthesis Using Text System consists of several integrated modules that work together to transform user-provided textual input into a complete musical composition. Each module performs a specific function in the music generation process and contributes to the overall efficiency, accuracy, and quality of the generated song. The User Interface Module acts as the communication layer between the user and the system. The Text Processing Module is responsible for analyzing the text provided by the user. The Melody Generation Module automatically generates musical notes and melodic sequences based on the processed text. The Vocal Synthesis Module converts textual lyrics into realistic human-like speech using Text-to-Speech technology. The Harmony and Rhythm Generation Module creates supporting musical elements such as chord progressions, bass lines, drum patterns, and rhythmic accompaniments. The Audio Processing and Mixing Module combines vocals, melodies, harmonies, and rhythm tracks into a single audio output.The Song Management Module handles the storage, retrieval, playback, and download of generated songs. The Database Management Module stores user information, generated song details, language selections, voice preferences, and system-related data.

1.Authentication Module:

1. Text Processing Module:

Handles user-provided text input and performs Natural Language Processing (NLP) operations to analyze lyrics, sentence structures, and language information before music generation.

I.2. Music Generation Module:

Generates melodies, harmonies, rhythm patterns, and instrumental arrangements automatically using Artificial Intelligence and MusicGen technologies.

II. 3. Vocal Synthesis Module:

Converts textual lyrics into natural-sounding vocals using Text-to-Speech technologies such as Bark AI, gTTS, and pyttsx3. It supports multilingual voice generation and different voice styles.

III. 4. Audio Processing Module:

Combines generated music and vocals, performs audio mixing, normalization, sound enhancement, and creates the final song output.

IV. 5. Song Management Module:

Handles song playback, audio storage, song retrieval, and MP3/WAV download functionality for generated musical compositions.

V. 6. Database Management Module:

Manages storage, retrieval, and maintenance of user information, generated song records, audio file details, and application settings.

.

4. EXPERIMENTAL RESULTS

The performance of the Automatic Music Composition and Vocal Synthesis Using Text System was evaluated based on music generation success rate, vocal synthesis accuracy, audio quality, and system response time. The experimental results demonstrate that the system efficiently converts textual input into complete musical compositions containing melodies, harmonies, vocals, and background music.

1) Music Generation Success Rate

Music Generation Success Rate measures the percentage of user inputs that are successfully converted into complete musical compositions.

Formula:

Music Generation Success Rate = (Number of Successfully Generated Songs / Total Number of Generation Requests) × 100

A higher success rate indicates better system reliability and performance.

2) Vocal Synthesis Accuracy

Vocal Synthesis Accuracy measures the correctness of generated speech with respect to the user-provided lyrics.

Formula:

Vocal Synthesis Accuracy = (Correctly Synthesized Words / Total Words) × 100

This metric evaluates pronunciation quality and speech generation effectiveness.

3) Audio Quality Evaluation

Audio quality is measured based on clarity, synchronization between vocals and music, and absence of audio distortions. Audio normalization and signal processing techniques improve the overall quality of generated songs.

4) Response Time

Response Time measures the amount of time required by the system to generate a complete song after receiving textual input.

Formula:

Response Time = Song Generation Completion Time − User Request Time

Lower response times indicate better system efficiency.

5) User Satisfaction Rate

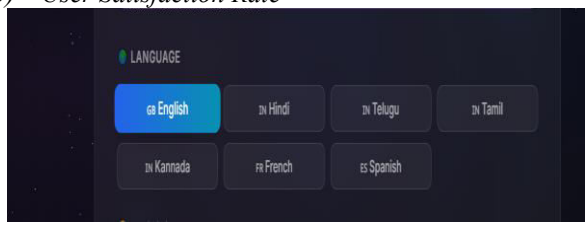

Fig2 This screen displays the language selection interface of the Automatic Music Composition and Vocal Synthesis Using Text System. The language selection module allows users to choose the preferred language for song generation and vocal synthesis before creating a song. This feature plays an important role in enabling multilingual music generation and improving user accessibility.

The system supports multiple languages including English, Hindi, Telugu, Tamil, Kannada, French, and Spanish. Users can simply select their desired language, and the system automatically processes the input text according to the selected language. The AI-powered Text-to-Speech module generates vocals in the chosen language while maintaining proper pronunciation, clarity, and natural speech quality.

The selected language is passed to the Natural Language Processing (NLP) module, which analyzes the text and prepares it for melody generation and vocal synthesis. This enables users from different linguistic backgrounds to create songs in their native language without any difficulty.

The multilingual capability of the system enhances flexibility and usability by allowing users to generate songs, poems, devotional music, and other musical content in different languages. This feature makes the application suitable for a wide range of users including students, content creators, musicians, educators, and entertainment enthusiasts

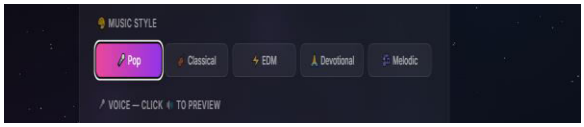

Fig3 This screen displays the music style selection interface of the Automatic Music Composition and Vocal Synthesis Using Text System. Before generating a song, users can select their preferred music style according to the mood and theme of their lyrics. The system provides multiple music styles including Pop, Classical, EDM, Devotional, and Melodic.

Each music style influences the melody, rhythm, harmony, tempo, and overall composition of the generated song. For example, Pop music produces modern and energetic compositions, Classical style generates soft and orchestral melodies, EDM creates electronic dance music patterns, Devotional style generates spiritual musical arrangements, and Melodic style focuses on smooth and pleasant tunes.

The selected music style is passed to the AI music generation module, which creates suitable instrumental arrangements and background music based on user preferences. This feature enhances personalization and allows users to generate songs that match their desired musical taste. The interface is designed to provide a simple and user-friendly experience while offering flexibility in music creation

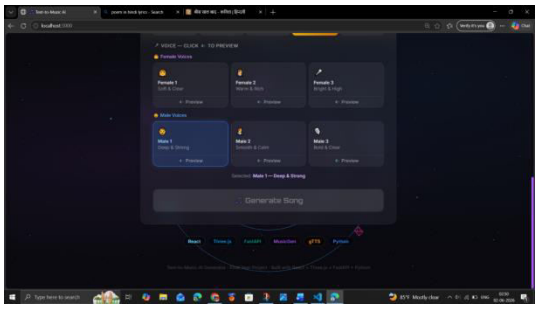

Fig4 This screen displays the voice selection interface of the Automatic Music Composition and Vocal Synthesis Using Text System. The system allows users to choose different voice options before generating a song. Multiple male and female voices are provided, each having unique voice characteristics such as tone, pitch, clarity, and speaking style.

Users can preview available voices and select the most suitable voice according to the song lyrics and music style. The selected voice is used by the Text-to-Speech and Vocal Synthesis modules to generate natural-sounding vocals for the final song. This feature helps users personalize the generated music and create songs with different vocal effects.

The interface provides an easy and interactive method for selecting voices while ensuring a smooth

user experience. By supporting multiple voice options, the system enhances creativity and allows users to generate songs with diverse vocal styles and expressions

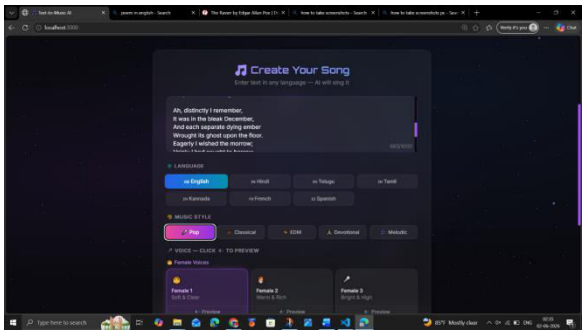

Fig5 This screen displays the text input and song configuration interface of the Automatic Music Composition and Vocal Synthesis Using Text System. Users can enter lyrics, poems, stories, or any textual content that will be converted into a complete musical composition using Artificial Intelligence.

The interface allows users to select their preferred language, music style, and voice type before generating the song. Multiple language options, music genres, and vocal styles are provided to personalize the generated output according to user preferences. The entered text is processed by Natural Language Processing techniques and forwarded to the AI music generation module.

After receiving the input, the system automatically generates melodies, harmonies, rhythms, and synthesized vocals to create a complete song. This page acts as the main configuration stage of the application and provides users with full control over the music generation process through a simple and interactive interface

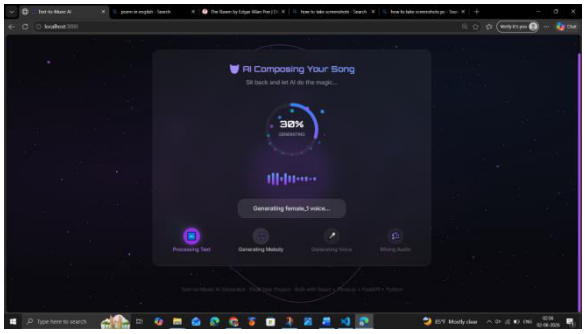

Fig 6: This screen displays the song generation progress interface of the Automatic Music Composition and Vocal Synthesis Using Text System. After receiving the user input, the Artificial Intelligence module begins processing the text and automatically generates the musical composition. The interface provides real-time progress updates and displays the current stage of song creation.

The system performs multiple operations including text processing, melody generation, vocal synthesis, and audio mixing. A progress indicator is displayed to show the completion status of the generation process, while the waveform animation visually represents ongoing audio creation activities.

The page keeps users informed about the current processing stage and ensures transparency during music generation. Once all stages are completed successfully, the system produces a fully composed song containing vocals, melody, harmony, and background music, which can then be played or downloaded as an MP3 file

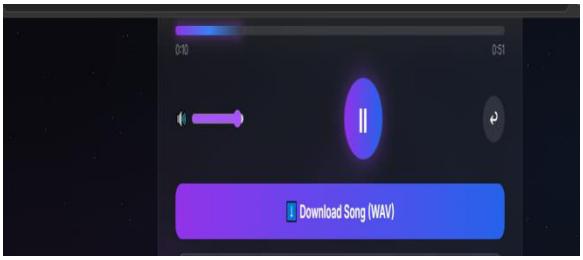

Fig 7: This screen displays the final output page of the Automatic Music Composition and Vocal Synthesis Using Text System after successful song generation. The generated song can be played directly through the integrated audio player, allowing users to listen to the AI-composed music with synthesized vocals.

The interface provides playback controls such as play, pause, volume adjustment, and progress tracking to enhance the listening experience. Users can review the generated composition before saving it to their device.

The system also provides a download option that allows users to save the generated song in audio format for future use. This feature enables users to easily share, store, and access their AI-generated musical compositions. The output page serves as the final stage of the music generation process, delivering a complete and downloadable song created from user-provided text

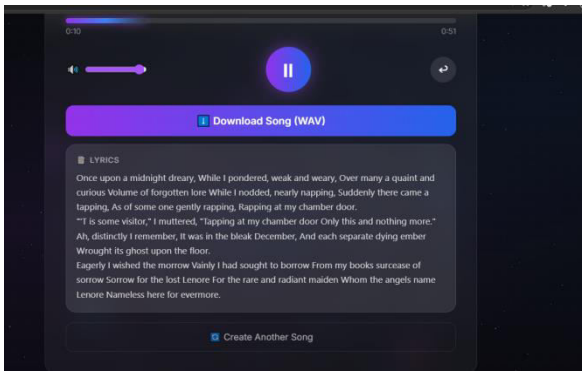

Fig 8: This screen displays the final output interface of the Automatic Music Composition and Vocal Synthesis Using Text System after successful song generation. The generated song can be played directly using the integrated audio player, allowing users to listen to the AI-generated music along with synthesized vocals.

The interface provides playback controls such as play, pause, volume adjustment, and progress tracking to improve the listening experience. It also includes a download option that enables users to save the generated song in WAV format for future use and sharing.

In addition to audio playback, the system displays the original lyrics used for song generation. This allows users to review the textual content while listening to the generated composition. The page also provides a "Create Another Song" option, enabling users to generate new musical compositions without restarting the application.

This output screen represents the final stage of the music generation process and delivers a complete song containing AI-generated melody, synthesized vocals, downloadable audio, and lyric visualization in a single user-friendly interface

5. CONCLUSION

• The Automatic Music Composition and Vocal Synthesis Using Text System was successfully designed, developed, and implemented to automatically generate complete musical compositions from user-provided textual input. The system provides an intelligent and efficient platform that combines Artificial Intelligence, Natural Language Processing, Music Generation, Vocal Synthesis, and Audio Processing technologies into a single integrated solution.

• The developed system allows users to enter lyrics, poems, stories, or any textual content and automatically converts them into fully composed songs. AI-based music generation techniques create melodies, harmonies, rhythm patterns, and instrumental arrangements, while Text-to-Speech technologies generate natural-sounding vocals in multiple languages

.• The integration of React.js frontend, FastAPI backend, MusicGen AI model, Bark AI, gTTS, pyttsx3, and audio processing libraries enabled smooth communication between all system components. Users can generate songs, listen to generated music, and download MP3 files through a simple and user-friendly interface.

• The system significantly reduces the complexity of music production and makes music creation accessible to users without requiring professional musical knowledge. The developed application provides a reliable, scalable, and intelligent solution for AI-powered music generation and vocal synthesis.

6. FUTURE SCOPE

• Although the current system performs automatic music composition and vocal synthesis effectively, several additional features can be implemented in future versions to improve system performance, creativity, and scalability.

Advanced Emotion-Based Music Generation

• Future versions can analyze user emotions from textual input and generate music that matches emotional states such as happiness, sadness, motivation, excitement, and relaxation.

Real-Time Music Generation

• The system can be enhanced to generate music in real time while users enter lyrics, providing a more interactive and dynamic experience.

Additional Music Genres

- More music styles such as Jazz, Hip-Hop, EDM, Classical, Folk, Rock, and Cinematic music can be integrated to provide greater creative flexibility.

Advanced AI Music Models

- More advanced generative AI models can be integrated to produce higher-quality melodies, harmonies, and instrumental compositions.

REFERENCES

- Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016.
- Simon Haykin, Neural Networks and Learning Machines, Pearson Education, 3rd Edition, 2009.
- Stuart Russell and Peter Norvig, Artificial Intelligence: A Modern Approach, Pearson Education, 4th Edition, 2020.
- Daniel Jurafsky and James H. Martin, Speech and Language Processing, Pearson Education.
- Christopher D. Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press.
- FastAPI Official Documentation, Available: https://fastapi.tiangolo.com/
- React Official Documentation, Available: https://react.dev/
- Python Official Documentation, Available: https://docs.python.org/3/
- NumPy Official Documentation, Available: https://numpy.org/doc/
- SciPy Official Documentation, Available: https://docs.scipy.org/
- PyDub Official Documentation, Available: https://github.com/jiaaro/pydub
- Google Text-to-Speech (gTTS) Documentation, Available: https://gtts.readthedocs.io/
- Pyttsx3 Documentation, Available: https://pyttsx3.readthedocs.io/
- Hugging Face Transformers Documentation, Available: https://huggingface.co/docs/transformers
- PyTorch Official Documentation, Available: https://pytorch.org/docs/
- MusicGen Documentation, Available: https://github.com/facebookresearch/audiocraft
- Bark AI Documentation, Available: https://github.com/suno-ai/bark
- Three.js Official Documentation, Available: https://threejs.org/docs/
- MDN Web Docs – HTML, CSS, JavaScript Documentation, Available: https://developer.mozilla.org/
- Various IEEE Research Papers, Journal Articles, and Technical Publications related to Artificial Intelligence, Music Generation, Natural Language Processing, Text-to-Speech Systems, and Generative AI Technologies..

Author Profiles

Mr. Y. Nagamalleswararao completed his Master of Technology (M.Tech) from JNTUK, M.Sc (IS) from ANU, and BCA from ANU. He has expertise in System Administration, Network Administration, and Oracle Administration. He is also a Web Developer and Python Developer. Currently, he is working as an Assistant Professor in the Department of MCA at SRK Institute of Technology, Enikepadu, NTR District. His areas of interest include Artificial Intelligence and Machine Learning.

Ms. K. Pavani is working as an Assistant Professor and Head of the Department of MCA at SRK

Institute of Technology, Vijayawada. She completed her M.Tech and MCA in Computer Science. She has 10 years of teaching experience at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. Her areas of interest include Machine Learning with Python and DBMS.

Mr. G.Ravi Teja is an MCA student in the Department of Computer Applications (MCA) at SRK Institute of Technology, Enikepadu, Vijayawada, NTR District. He completed his degree in B.Sc. (Computers) from Andhra Loyola College, Vijayawada. His areas of interest include Java and Machine Learning with Python.