International Journal of Creative and Open Research in Engineering and Management
ISSN: 3108-1754 (Online)
Volume 02 Issue 10 October-2026 | Impact Factor: 3.5



AI-Based Voice to Music Generation System

ABDUL RAZZAQ
Department of CSE
VTU University
Belagavi, India
abdulrazzak33824@gmail.com

ARUN KUMAR KH
Department of CSE
VTU University
Belagavi, India
arunkumarkh187676@gmail.com

BHASKAR V
Department of CSE
VTU University
Belagavi, India
bv254224@gmail.com

LOHITH C N
Department of CSE
VTU University
Belagavi, India
lohi18th2005cn@gmail.com

Guide : Mounashree HS

Asst Prof of CSE department VTU University Belagavi,India

hsmounashree2711@gmail.com

Abstract—

The rapid advancement of Artificial Intelligence (AI), Machine Learning (ML), Natural Language Processing (NLP), and generative models has created new opportunities for automating creative processes, particularly in music composition and audio production. Traditional music creation often requires musical knowledge, instrument-playing skills, professional audio-editing tools, and considerable time. These requirements can make music production challenging for beginners and individuals without formal musical training. To address these challenges, this research proposes an **AI-Based Voice-to-Music Generation System** that enables users to create musical compositions through natural voice commands.

The proposed system aims to integrate speech recognition, natural language processing, audio signal processing, and AI-based music generation into a unified framework. Users can provide voice commands through a microphone or upload audio recordings describing their musical preferences. The system processes the input using speech recognition techniques to convert speech into text. Natural Language Processing is then applied to identify relevant musical characteristics, including genre, mood, tempo, and instrument preferences. Based on the extracted information, an appropriate generative music model produces a musical composition corresponding to the user's instructions.

The generated audio is subsequently processed and made available for playback, saving, and downloading through an interactive user interface. The effectiveness of the proposed system will be investigated through experiments evaluating speech recognition accuracy, musical command interpretation, music generation quality, processing time, and user satisfaction.

Keywords— Artificial Intelligence, Voice-to-Music Generation, Natural Language Processing, Speech Recognition, Machine Learning, Deep Learning, Generative Music Models, Audio Signal Processing, Music Composition.

How to Cite this Article:

RAZZAQ, A., KH, A. K., V, B. & N, L. C. (2026). AI-Based Voice to Music Generation System. International Journal of Creative and Open Research in Engineering and Management, 02(10), 1-9. https://doi.org/10.55041/ijcope.v2i10.179

License:

This article is published under the terms of the Creative Commons Attribution 4.0 International License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

© The Author(s). Published by International Journal of Creative and Open Research in Engineering and Management.







https://doi.org/10.55041/ijcope.v2i10.179



I. INTRODUCTION

Artificial Intelligence (AI) has emerged as an important technology in various fields, including healthcare, education, communication, entertainment, and creative industries. Recent advancements in Machine Learning (ML), Deep Learning (DL), Natural Language Processing (NLP), and generative models have enabled computers to perform tasks that traditionally require human intelligence and creativity. One such emerging application is AI-based music generation, which focuses on creating musical compositions automatically based on user instructions, musical patterns, or audio inputs. These developments provide new possibilities for simplifying music production and making creative tools accessible to a wider range of users.

Music composition is traditionally a creative process that requires knowledge of musical theory, rhythm, melody, harmony, and instrument selection. Professional music production may also involve digital audio workstations, recording equipment, sound-editing software, and considerable manual effort. Although these tools provide extensive creative control, they can present challenges for beginners and individuals without formal musical training. Therefore, developing intelligent systems that simplify the music creation process has become an important area of research.

The proposed AI-Based Voice-to-Music Generation System aims to provide an interactive approach to music composition by allowing users to describe their musical preferences through natural voice commands. Instead of manually selecting instruments, arranging musical notes, or editing audio tracks, users can provide spoken instructions describing the type of music they wish to generate. These instructions may include characteristics such as musical genre, emotional mood, tempo, and preferred instruments. The system interprets these commands and uses an AI-based music generation model to produce a corresponding musical composition.

The proposed framework integrates several technologies to achieve this objective. Speech recognition is used to convert spoken input into textual information. Natural Language Processing techniques analyse the recognized text and identify the user's intended musical preferences. representation that can guide the music generation process.

II .LITERATURE REVIEW

1. Agostinelli et al. (2023) – MusicLM: Generating Music from Text

The authors proposed MusicLM, an artificial intelligence model that generates music from natural language descriptions. The study demonstrated how text descriptions can be used to guide the generation of high-quality audio with musical consistency. The model can capture characteristics such as musical style, mood, and instruments from textual prompts. However, the system relies on text-based input, and integrating direct voice commands requires additional speech recognition and language-processing techniques.

2. Copet et al. (2023) – Simple and Controllable Music Generation

The researchers introduced MusicGen, a music generation model that creates music using text descriptions and other conditioning inputs. The study demonstrated the use of language models and audio token representations to generate music according to user instructions. The model supports controllable music generation and provides a foundation for developing interactive music applications. However, converting spoken commands into suitable text prompts and ensuring accurate interpretation of multilingual instructions remain additional challenges for a voice-controlled system.

3. Liu et al. (2023) – AudioLDM 2: Learning Holistic Audio Generation with Self-Supervised Pretraining

The authors presented AudioLDM 2, a general-purpose audio generation framework capable of generating different types of audio from text and other conditioning information. The approach uses self-supervised audio representations and language-related modelling techniques to connect input descriptions with generated audio. The study demonstrated the potential of using a unified framework for multiple audio-generation tasks. However, adapting the model specifically for music composition controlled by spoken instructions requires additional processing to identify musical preferences and translate them into effective generation prompts.



III. METHODOLOGY

The proposed AI-Based Voice-to-Music Generation System follows an artificial intelligence-based approach that integrates voice input collection, speech recognition, natural language processing, prompt generation, music generation, and performance evaluation. The system is designed to analyse spoken instructions, identify musical preferences, and generate corresponding musical compositions. It follows a structured workflow in which voice inputs are processed before being supplied to an AI-based music generation model.

The integration of these components provides an automated platform for voice-controlled music creation. The methodology begins with voice input collection and preprocessing. Voice commands are collected through a microphone or uploaded audio recordings containing instructions about musical characteristics such as genre, mood, tempo, and instrument preferences. The system is intended to support English and Hindi voice inputs. The collected audio is processed using a suitable speech recognition model to convert spoken instructions into text. The recognized text is then analysed using Natural Language Processing (NLP) techniques to identify the user's intended musical preferences. The extracted information is organized into a structured representation that serves as the basis for generating an appropriate music prompt.

The prompt generation and music synthesis process forms the core of the proposed system. The identified musical characteristics are converted into descriptive text prompts and supplied to a suitable AI-based music generation model. Generative models use learned musical and audio patterns to produce compositions that attempt to match the requested style, mood, and other musical attributes. The generated audio is subsequently processed using audio-processing libraries to support playback, storage, and downloading. A user-friendly interface can be developed using Streamlit or web technologies, with Python supporting the speech-processing and music-generation components.

IV. RESULTS AND DISCUSSION

The proposed AI-Based Voice-to-Music Generation System aims to provide an automated approach to music composition by integrating speech recognition, Natural Language Processing (NLP), and AI-based music generation. The system is designed to convert spoken instructions into musical compositions based on user preferences, including genre, mood, tempo, and instrument selection.

1 Home Interface

The AI-Based Voice-to-Music Generation System is an intelligent application designed to simplify music creation using Artificial Intelligence. The home page provides a user-friendly interface through which users can access the system's main features. Users can provide voice commands in English or Hindi to describe their desired musical composition. The system processes the voice input using speech recognition and Natural Language Processing techniques.

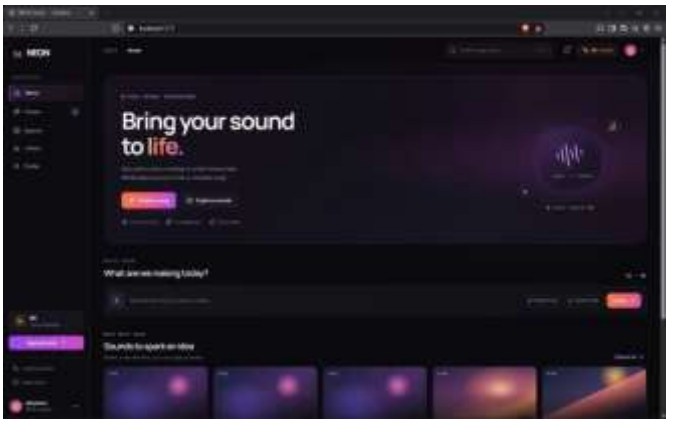


Fig 1. Home Interface

2. Login Interface

The login interface of the AI-Based Voice-to-Music Generation System provides secure access to registered users. It allows users to enter their email address or username and password to access the application. The interface is designed with a simple and user-friendly layout to provide a smooth authentication experience. New users can register by providing the required details through the sign-up option.

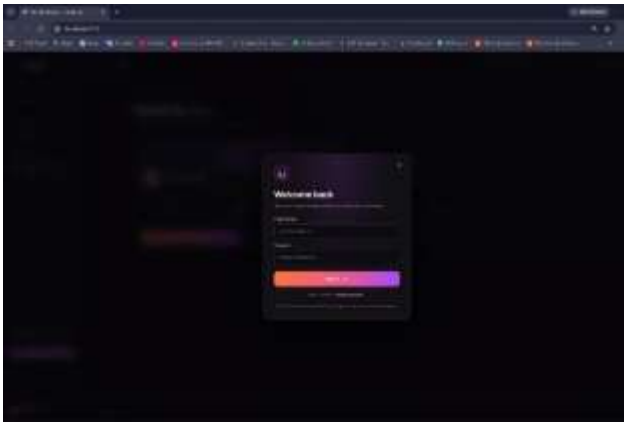


Fig 2. Login interface

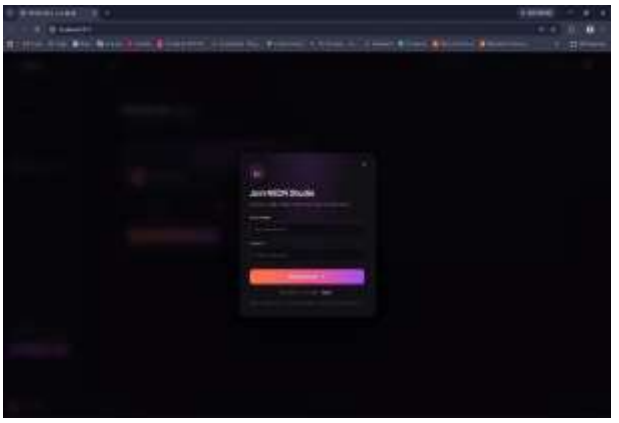


Fig.3 Registration Interface

3. Registration

The registration page of the proposed AI-Based Voice-to-Music Generation System is designed to enable new users to create personal accounts and gain access to the application's features. It serves as an important component of the user authentication process and provides a structured interface for collecting user details.

The registration form includes fields such as full name, email address, password, and confirm password. Users are required to enter valid information to complete the registration process successfully.

The interface incorporates input validation to identify missing fields, invalid email formats, weak passwords, and mismatched passwords. Appropriate validation messages can be displayed to help users correct errors before submitting the form. For security, passwords should be stored using a suitable password-hashing algorithm rather than as plain text. After successful registration, the system can save the user's account information in a database and redirect the user to the login page.

Once registered, users can log in using their credentials and access the application's main dashboard. From the dashboard, they can provide voice commands, select musical preferences, generate music using the integrated AI model, and play or download the resulting audio.

4. Creat Interface

The Create Song interface of the proposed AI-Based Voice-to-Music Generation System allows users to generate musical compositions based on their preferences and voice commands. It provides an interactive platform where users can describe the type of music they want to create, including genre, mood, tempo, and instrument selection.

Users can enter their requirements through voice input or text input, depending on the available system features. The voice input is processed using speech recognition to convert spoken instructions into text, while Natural Language Processing (NLP) techniques identify the relevant musical characteristics.

These characteristics are converted into a suitable prompt for the AI-based music generation model. After processing the request, the model generates an audio composition according to the provided instructions.

The interface displays the generation status and provides audio playback controls to listen to the generated song. Users can also save or download the generated audio for future use. The interface is designed to be simple and user-friendly, enabling beginners and individuals without professional musical knowledge to explore AI-powered music creation.

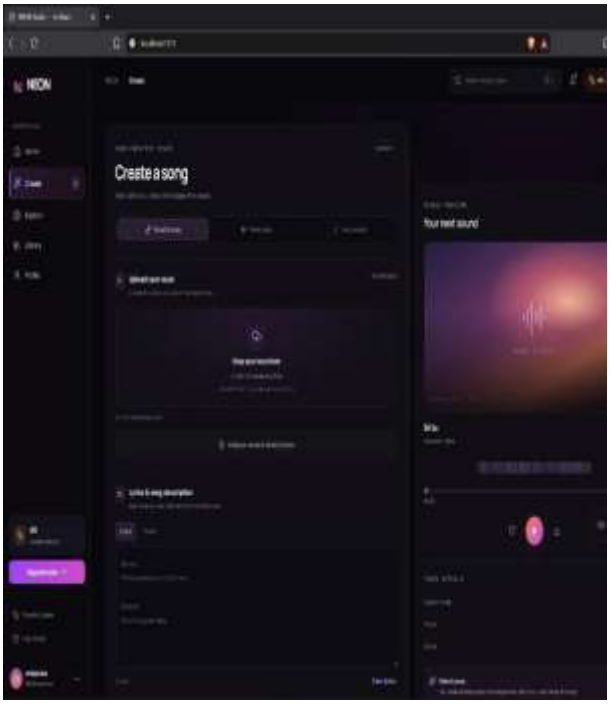


Fig.4 Creat interface

5. Result Interface

The Song Result page of the proposed AI-Based Voice-to-Music Generation System displays the musical composition generated from the user's voice commands or text prompts. After the AI model completes the music generation process, the resulting audio is presented through a simple and interactive interface. The page displays the song title, generation status, and available audio playback controls, allowing users to play, pause, and listen to the generated composition. Users can review the generated music to determine whether it matches their requested genre, mood, tempo, and instrument preferences. The interface can also provide options to download the audio file, save the composition for future access, or generate another song using different instructions. If the generated music does not meet the user's expectations, the user can return to the Create Song page and modify the input prompt. The system may also display relevant information such as audio duration and file format. The Song Result page provides convenient access to generated compositions and completes the music creation workflow, from voice input and prompt processing to audio generation and playback. Its primary objective is to deliver an accessible and user-friendly experience for exploring AI-generated music.



Fig.5 Result Interface

V. CONCLUSION

The proposed AI-Based Voice-to-Music Generation System presents an intelligent approach to simplifying music composition through the integration of Artificial Intelligence (AI), speech recognition, Natural Language Processing (NLP), and generative music models. Traditional music production often requires musical knowledge, instrument selection, and specialized editing tools, which may create difficulties for beginners. The proposed system aims to overcome these challenges by allowing users to generate music through natural voice commands without requiring advanced musical skills.

The system is designed to accept voice inputs in English and Hindi and identify musical preferences such as genre, mood, tempo, and instrument selection. Speech recognition converts the spoken instructions into text, while NLP techniques analyse the input and extract the relevant musical characteristics.



VI. REFERENCES

[1] A. Agostinelli et al., "MusicLM: Generating Music From Text," *arXiv preprint arXiv:2301.11325*, 2023.

[2] J. Copet et al., "Simple and Controllable Music Generation," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.

[3] H. Liu et al., "AudioLDM 2: Learning Holistic Audio Generation with Self-Supervised Pretraining," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 2871–2883, 2024.

[4] Y. Zhao, M. Yang, Y. Lin, X. Zhang, F. Shi, Z. Wang, J. Ding, and H. Ning, "AI-Enabled Text-to-Music Generation: A Comprehensive Review of Methods, Frameworks, and Future Directions," *Electronics*, vol. 14, no. 6, article 1197, 2025.

[5] J. Copet, F. Kreuk, I. Gat, T. Remez, D. Kant, G. Synnaeve, Y. Adi, and A. Défossez, "Simple and Controllable Music Generation," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.

[6] H. Liu, Y. Yuan, X. Liu, X. Mei, Q. Kong, Q. Tian, Y. Wang, W. Wang, Y. Wang, and M. D. Plumbley, "AudioLDM 2: Learning Holistic Audio Generation with Self-Supervised Pretraining," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 32, pp. 2871–2883, 2024.