IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14580

Advanced Multimodal Podcast Orchestration Framework

Prajwal Ullas Naik [1], Sanket [2], Rohith B M [3], Sumanth U S [4], Mrs. Tejashree V [5]

Students, Information Science and Engineering, SJB Institute of Technology, Bengaluru, India [1-4]

Assistant Professor, Information Science and Engineering, SJB Institute of Technology, Bengaluru, India [5]

Abstract: In today's rapidly evolving digital landscape, podcast creators face significant challenges in content management, accessibility, and audience engagement. Traditional podcasting platforms often lack automation, requiring manual efforts for editing, transcription, and distribution. Additionally, discoverability issues and limited collaboration features hinder creators from reaching wider audiences efficiently. Our project introduces an advanced multimodal podcast orchestration framework, leveraging AI-driven tools and cloud-based automation to streamline podcast production. By integrating modern technologies such as Next.js, React.js, Convex, and OpenAI, our system enhances content management with automated `transcription, AI-powered content summarization, and intelligent recommendations. The platform offers real-time collaboration tools, allowing podcasters and editors to work seamlessly. Secure user authentication is implemented via Clerk, ensuring data integrity and controlled access. Unlike traditional solutions, our system provides a scalable and adaptive infrastructure, enabling smooth performance under high workloads. Through automated editing, voice processing, and intelligent tagging, our framework reduces the manual workload on creators, improving efficiency and content quality. The integration of machine learning models enhances content discoverability, making personalized recommendations based on listener behavior. By providing a unified solution for hosting, organization, and distribution, this system significantly simplifies podcast production. The proposed framework ensures accessibility, multi-device compatibility, and AI-enhanced automation, addressing current limitations in podcast management. Future advancements may include monetization tools, real-time translation, and deeper AI integration to further enhance the ecosystem. By leveraging cutting-edge technologies, this project aims to revolutionize podcasting, empowering creators with tools that optimize workflows, maximize audience reach, and redefine digital audio experiences.

Keywords: Podcast Management, AI Automation, Cloud-Based Framework, Content Discovery, Machine Learning, Audio Processing, Real-Time Collaboration, User Authentication.

1. INTRODUCTION

In today's interconnected digital landscape, podcast creation and management have become increasingly complex, requiring creators to juggle multiple tools for recording, editing, transcription, and distribution. Traditional podcasting workflows rely on manual efforts, leading to inefficiencies and limiting scalability. Additionally, challenges such as content discoverability, accessibility for diverse audiences, and real-time collaboration hinder creators from maximizing their reach and engagement. While many professional podcasters invest in high-end software for content production and management, these solutions are often expensive, require extensive technical knowledge, and lack seamless integration across the entire podcasting workflow. Small and independent creators, in particular, face difficulties in adopting these tools due to cost and complexity, restricting their ability to compete with larger media houses.

One effective approach to addressing these challenges is leveraging AI-driven automation and cloud-based orchestration to streamline the podcasting process. By integrating modern technologies such as Next.js, React.js, Convex, and OpenAI, it is possible to enhance workflow efficiency, automate repetitive tasks, and improve content accessibility. Additionally, implementing intelligent recommendation systems and automated transcription services can help podcasters optimize content discoverability while ensuring inclusivity for all audience groups.

This research explores the feasibility of an AI-powered, multimodal podcast orchestration framework that integrates automated editing, real-time collaboration, and content management tools. The proposed system aims to simplify podcast production, improve accessibility, and enhance scalability while minimizing reliance on multiple third-party applications. By providing a unified, cloud-based solution, podcast creators can achieve a cost-effective, efficient, and future-ready approach to content management.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025
DOI: 10.17148/IJARCCE.2025.14580

2. LITERATURE SURVEY

As podcasting technology evolves, optimizing podcast creation and management is crucial for enhancing accessibility, automation, and audience engagement. Traditional podcasting platforms rely on manual editing, metadata tagging, and distribution processes, which often lack efficiency and scalability. Modern approaches leverage artificial intelligence (AI), machine learning (ML), and cloud computing to streamline workflows and enhance user experience. This survey examines existing methodologies, their strengths, and emerging trends, forming the basis for an AI-powered multimodal podcast orchestration framework.

Kevin Goldberg et al. [1] conducted an extensive study on AI-driven automation in podcasting. The authors detail a range of techniques, including speech-to-text conversion, intelligent content tagging, and automated editing. Machine learning models trained on podcast metadata, audience behavior, and language processing algorithms help generate personalized recommendations, automated summaries, and improved transcription accuracy, enhancing content discoverability. The study highlights challenges such as audio quality enhancement, reducing transcription errors, automating noise removal, and improving multilingual support to ensure a seamless podcasting experience for diverse audiences. The researchers advocate for integrating deep learning, cloud-based automation, and AI-assisted voice modulation to address these challenges and improve workflow efficiency.

Daniel J. Lewis et al. [2] explored AI-based content discovery and recommendation systems. The study emphasizes the role of natural language processing (NLP), collaborative filtering, and sentiment analysis in optimizing content suggestions based on listener preferences. The authors provide case studies on how platforms like Spotify, Apple Podcasts, and Google Podcasts use ML algorithms to analyze listening habits, genre preferences, and engagement metrics, thereby improving personalized recommendations.

E. M. Schmidt and Y. Kim [3] investigated automated audio editing techniques and deep learning-based speech processing in podcast production. The paper reviews AI-powered noise reduction, speech enhancement, real-time pitch correction, and emotion-based audio modulation. The authors analyze the effectiveness of generative adversarial networks (GANs), deep neural networks (DNNs), and unsupervised learning models in improving audio clarity while reducing manual effort. The study concludes that integrating AI into audio processing pipelines and leveraging cloud-based computational resources significantly enhances production efficiency and ensures high-quality podcast output without requiring extensive manual intervention.

James Cridland et al. [4] studied real-time collaboration tools and cloud-based synchronization techniques for podcast production. The research explores remote recording solutions, live content editing, automated speaker identification, and AI-assisted script generation. The findings highlight the benefits of AI-driven automation in transcription accuracy, multi-speaker differentiation, content structuring, and seamless collaboration. The authors emphasize the importance of integrating secure authentication systems, blockchain-based content verification, and distributed cloud infrastructure to enhance accessibility and usability for podcast creators, ensuring a highly scalable and collaborative environment.

Naveed Ahmed et al., [4] explores the integration of machine learning (ML) and deep learning (DL) techniques into Software-Defined Networking (SDN)-based Network Intrusion Detection Systems (NIDS). SDN enables centralized network control but also introduces vulnerabilities, necessitating advanced security measures. The study reviews existing ML- and DL-based NIDS approaches, highlighting their effectiveness in identifying security threats while addressing challenges like high false alarm rates and computational complexity. Finally, it outlines future research directions, aiming to enhance detection accuracy and mitigate evolving cyber threats in SDN environments.

3. METHODOLOGY

In this paper, the AI-driven multimodal podcast orchestration framework is focused on enhancing podcast creation and management by automating transcription, editing, tagging, and distribution. When a creator uploads an audio file, the system processes the content using AI-powered transcription models, metadata tagging, and content structuring algorithms. It then analyses the data, generates transcripts, applies noise reduction, and optimizes audio clarity before finalizing the episode.

If an audio file requires modifications, the system automatically suggests enhancements such as voice balancing, adaptive filtering, and AI-assisted editing to refine the quality. This ensures a streamlined production process while minimizing manual effort. The system is lightweight, scalable, and integrates seamlessly with existing podcast platforms, making it a reliable and efficient solution for content creators aiming to automate and simplify podcast production.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14580

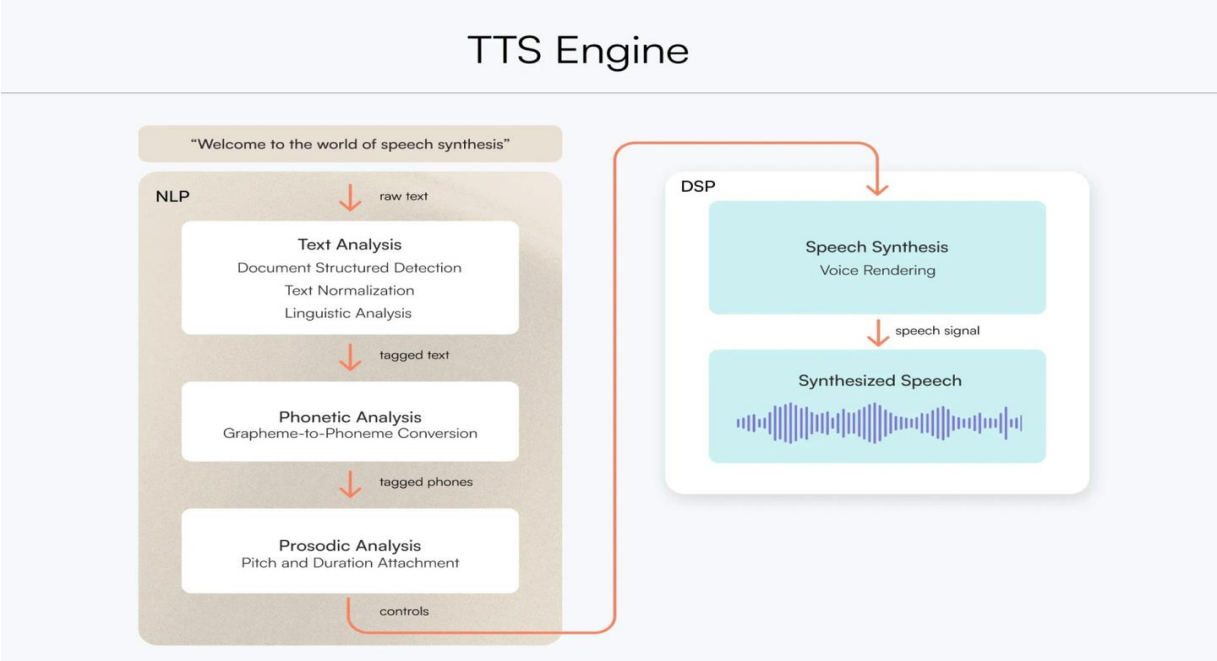

Fig.1: Data flow diagram (Text to Speech)

The podcast orchestration system is designed to enhance podcast production by automating editing, transcription, and publishing tasks. The process begins with the system receiving an audio file or script as input, which is then processed using a cloud-based automation framework. This framework systematically applies AI-driven audio enhancements, such as automatic segmentation, voice recognition, and keyword extraction. The system also integrates third-party speech-processing APIs to generate accurate transcripts, detect speaker changes, and analyze engagement metrics.

Upon receiving processed results, the system aggregates content metadata and optimizes podcast structuring for improved discoverability across multiple streaming platforms. It analyzes audio characteristics, episode themes, and listener preferences to enhance categorization, making it easier for audiences to find relevant content. However, if further refinement is needed, automated enhancement tools apply advanced noise reduction algorithms, dynamic audio balancing, volume normalization, and adaptive filtering to ensure a polished final product. These adjustments not only improve clarity and listener experience but also reduce manual post-production efforts**.** Furthermore, the system supports real-time content tagging and metadata generation, automatically extracting key topics, sentiment analysis, and engagement trends to ensure higher search engine visibility, enhanced recommendations, and better audience engagement. By automating the entire podcast workflow, from recording to distribution, the framework provides a scalable, efficient, and cost-effective solution for content creators seeking high-quality production with minimal effort, allowing them to focus on creativity while AI handles tedious technical tasks.

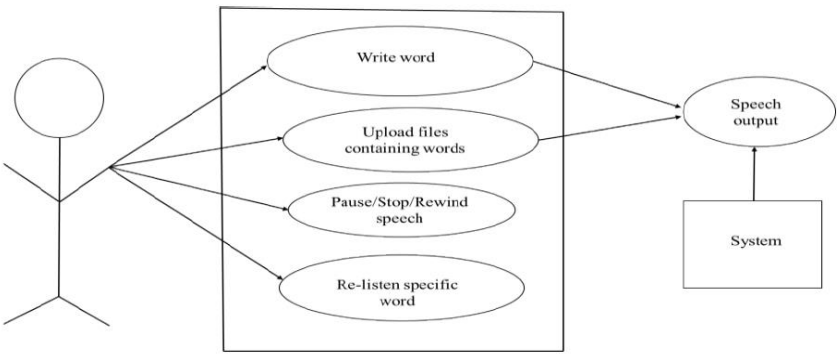

Fig.2: Use Case Diagram (Speech Generation)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14580

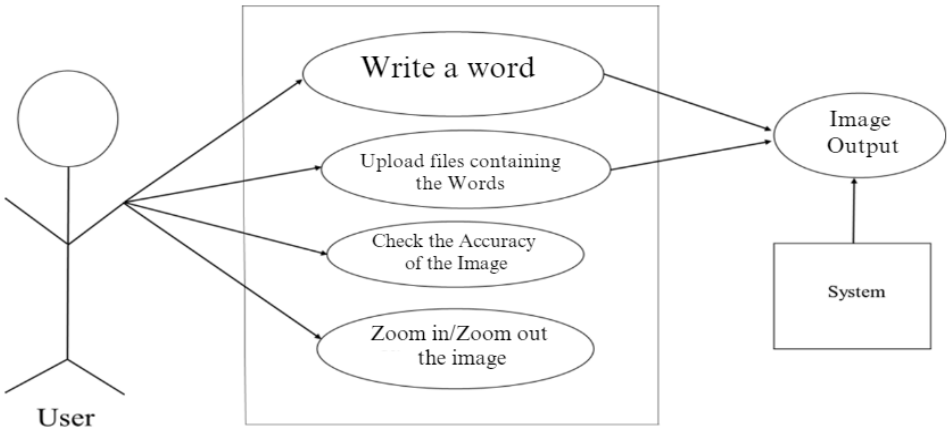

Fig.3: Use Case Diagram (Image Generation)

Independent podcasters and small production teams often struggle with complex post-production processes, making it challenging to efficiently create and manage content. This AI-powered framework offers a streamlined, automated approach that enhances podcast production without requiring extensive human intervention. By continuously analyzing and processing audio content, it eliminates manual transcription and metadata tagging, reducing production time and improving overall efficiency.

The automated workflow ensures that podcasts are organized, edited, and tagged without requiring extensive post-production labor, allowing creators to focus more on content strategy and engagement. Additionally, the local database stores optimized episodes and transcripts, further improving performance by reducing redundant manual effort when making revisions or updates.

Beyond individual content creation, this podcast framework also serves as a collaborative production tool where multiple creators, editors, and managers can share project data and manage workflows efficiently. When an episode is edited and finalized, it can be shared across teams, enabling seamless coordination between contributors.

By leveraging shared AI-powered editing tools, real-time audio optimization, and automated metadata tagging, independent podcasters can enhance production quality without the cost of hiring professional post-production teams. This scalable and collaborative approach strengthens the podcast industry by providing creators with a powerful, AI-driven solution for efficient, accessible, and high-quality content creation.

4. FUTURE SCOPE

The future development of the AI-driven multimodal podcast orchestration framework aims to enhance its efficiency, scalability, and adaptability in addressing evolving challenges in podcast production, content management, and audience engagement. Integrating AI-powered audio processing will enable the system to analyse speech patterns, detect emotion-based tone variations, and enhance voice clarity with greater accuracy Additionally, automated editing and content structuring mechanisms can be implemented to dynamically refine audio quality, adjust background music levels, and segment episodes for better listener retention. Expanding data sources by incorporating real-time audience feedback, AI-driven content recommendations, and multi-platform distribution analytics will further improve podcast visibility and engagement Expanding data sources by incorporating real-time audience feedback, AI-driven content recommendations, and multi-platform distribution analytics will further improve podcast visibility and engagement.

Performance optimization through local caching of frequently accessed podcast content, adaptive encoding for various devices, and AI-driven predictive content enhancement will reduce processing time, ensuring faster publishing and seamless user experience. Furthermore, developing an interactive podcast analytics dashboard will provide real-time insights into listener engagement, retention rates, and audience demographics, allowing creators to make data-driven content decisions. Additionally, automated multi-language transcription and AI-powered translation services will expand accessibility to global audiences, making podcasts more inclusive and discoverable across different regions.

Lastly, mobile and cloud integration will extend the system's capabilities, enabling seamless podcast production across devices and direct cloud-based publishing. This will allow podcasters to create, edit, and distribute content from anywhere, ensuring a scalable, cost-effective, and automated solution tailored for independent creators and podcast networks alike. By leveraging AI, automation, and cloud technologies, the proposed framework will redefine modern podcasting workflows, ensuring efficiency, accessibility, and audience growth in an evolving digital landscape.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2025

DOI: 10.17148/IJARCCE.2025.14580

5. CONCLUSION

In this research, we proposed an AI-driven multimodal podcast orchestration framework, an intelligent and automated solution designed to enhance podcast creation, management, and distribution by integrating AI-powered transcription, editing, tagging, and recommendation systems. By leveraging advanced speech-to-text models, cloud-based automation, and real-time content optimization, the framework ensures efficient podcast production and seamless content delivery across multiple platforms. The system not only automates audio enhancement and metadata generation but also fosters a collaborative production environment where creators can streamline workflows and improve audience engagement.

Its automated nature reduces manual intervention, minimizing production time and allowing podcasters to focus on content strategy rather than tedious post-production tasks. With future enhancements such as AI-driven voice modulation, personalized audio recommendations, and multi-language transcription, the proposed framework has the potential to evolve into a comprehensive, scalable, and industry-standard podcasting solution.

By enabling independent creators, podcast studios, and media networks to automate production, optimize discoverability, and enhance audience accessibility, this system contributes to a more efficient, engaging, and technology-driven podcasting ecosystem. Ensuring cost-effectiveness, scalability, and seamless integration with cloud platforms, this research lays the foundation for the future of AI-powered podcast production, making it an essential tool for modern content creators in the evolving digital lands.