International Journal Of Engineering Research And Development
e- ISSN: 2278-067X, p-ISSN: 2278-800X, www.ijerd.com
Volume 21, Issue 4 (April 2025), PP 167-170

Tamcast Android App

Anushka Bukkawar, Keyur Patil, Shravani Dore, Aditya Dhoble, Shweta S. Meshram*

[1]*Department of Information Technology, Faculty of Engineering*
Prof. Ram meghe college of Engineering and management, Badnera, Amravati, INDIA
Corresponding Author: Prof. Shweta S. Meshram

ABSTRACT
TamCast is a next-generation Android application designed to revolutionize podcast creation by seamlessly blending Text-to-Speech (TTS) technology with multimedia integration. Built using Java, XML, and Firebase, the app empowers users to generate high-quality podcasts by converting text into natural-sounding speech and synchronizing it with selected video content.
The application leverages Firebase Authentication, Realtime Database, and Cloud Storage to ensure secure access, real-time data management, and seamless media storage. A unique Podcast Library feature enriches the user experience by aggregating external podcasts through WebView, expanding content accessibility beyond self-created materials.
TamCast further enhances user engagement with an interactive feedback system, allowing creators to refine and improve their content. The My Podcast section offers a structured repository for storing, organizing, and managing all user-generated podcasts efficiently.
By integrating AI-driven speech synthesis, a web-powered content library, and a secure cloud-based infrastructure, TamCast redefines podcast creation, making it more intuitive, dynamic, and accessible to creators of all experience levels.

Key Words: *Android podcast app, Text-to-Speech (TTS), Firebase integration, Java Android development, podcast creation, video integration, WebView podcast scraping, cloud-based media storage, user-generated content, audio-to-video synthesis, AI-powered podcasting*

Date of Submission: 08-04-2025 Date of acceptance: 19-04-2025

I. INTRODUCTION

In the digital age, podcasts have emerged as a powerful medium for content creation, storytelling, and information sharing. However, traditional podcast production often involves complex recording setups, editing tools, and technical expertise, making it challenging for beginners and casual creators to participate. TamCast addresses this gap by offering an intuitive and efficient solution that simplifies podcast creation through Text-to-Speech (TTS) technology and video integration.

Developed using Java, XML, and Firebase, TamCast enables users to convert text into speech and seamlessly integrate it with a selected video, eliminating the need for professional voice recording. The app leverages Firebase Authentication, Real-time Database, and Cloud Storage to ensure secure access, real-time data management, and media storage.

Beyond personal content creation, TamCast enhances user experience by incorporating a Podcast Library, which scrapes external podcast data using WebView, giving users access to a diverse range of podcasts. Additionally, the app includes a feedback system to encourage user interaction and content refinement. All user-generated podcasts are systematically stored in the My Podcast section, allowing easy access and management.

With its AI-driven speech synthesis, cloud-based infrastructure, and web-integrated content discovery, TamCast redefines podcasting by making it more accessible, engaging, and effortless for creators of all skill levels. Whether for professional storytelling, educational content, or casual expression, TamCast empowers users to bring their ideas to life with minimal effort.

II. LITERATURE SURVEY

The advancements in Text-to-Speech (TTS) technology have revolutionized automated content creation, significantly enhancing accessibility and engagement in digital media. Heiga Zen (2016) introduced WaveNet, a deep learning-based TTS model that improved speech synthesis by generating more natural and human-like voices. Later, Jonathan Shen et al. (2018) refined TTS by incorporating Mel spectrogram predictions, which enhanced the expressiveness and intonation of AI-generated speech. These developments laid the foundation for AI-driven narration, making it an essential component of modern multimedia applications. TamCast leverages these advancements to provide users with an immersive and high-quality podcasting experience, eliminating the need for manual voice recording.

In the domain of content accessibility and aggregation, Christopher Olston (2010) explored web scraping methodologies for large-scale data extraction, emphasizing structured and real-time updates. This research supports TamCast's WebView-based Podcast Library, which enables users to explore external podcasts seamlessly without manual searches. By integrating real-time content aggregation, TamCast enhances content discovery and broadens user engagement.

Efficient media storage and security are critical for managing podcast data. Jeffrey Dean et al. (2012) analyzed cloud-based distributed storage systems, highlighting their ability to provide real-time synchronization, scalability, and secure authentication. TamCast adopts these principles by integrating Firebase Real-time Database and Cloud Storage, ensuring seamless podcast management while safeguarding user data.

Beyond technological infrastructure, user engagement and retention play a crucial role in digital platform success. Nir Eyal (2014) explored strategies for building habit-forming products, emphasizing interactive UI elements and feedback systems to enhance user experience. Inspired by this, TamCast incorporates an interactive feedback mechanism, allowing users to rate and improve podcast content dynamically, and fostering continuous engagement.

By combining AI-powered TTS synthesis, automated content aggregation, secure cloud storage, and user engagement strategies, TamCast introduces an innovative approach to podcasting. Unlike traditional platforms that rely on manual narration and hosting, TamCast automates the process, making podcast creation more accessible, efficient, and engaging. This literature survey underscores the technological and theoretical foundations that drive TamCast's development, positioning it as a comprehensive and user-friendly podcasting solution in the evolving digital audio landscape.

III. METHDOLOGY

The development of TamCast follows a structured methodology to ensure seamless podcast creation, efficient system integration, and enhanced user engagement. The process begins with requirement analysis and system design, where core features such as Text-to-Speech (TTS) conversion, video integration, Firebase storage, and podcast library aggregation are identified. The system architecture is designed to support real-time processing, authentication, and cloud storage, ensuring smooth functionality and scalability. The frontend and user interface development phase involves designing an intuitive UI using XML for layouts and Java for interactive functionalities in Android Studio. The interface allows users to input text, select a video, and generate podcasts effortlessly. Additionally, a Podcast Library is integrated using WebView, enabling users to explore external podcast content without manual searches.

For Text-to-Speech (TTS) integration, Google TTS API is utilized to convert user-inputted text into high-quality, natural-sounding speech, which is then merged with the selected video for an engaging podcast experience. The Firebase backend implementation ensures secure and scalable operations using Firebase Authentication, Real-time Database, and Cloud Storage. Authentication manages user registration and login, Real-time Database stores podcast metadata and user interactions, while Cloud Storage securely handles media files, user-generated content, and AI-generated audio. To enhance content accessibility, TamCast incorporates a WebView-based Podcast Library, where real-time web scraping is implemented to extract and update podcast content dynamically. This feature allows users to discover and stream external podcasts seamlessly.

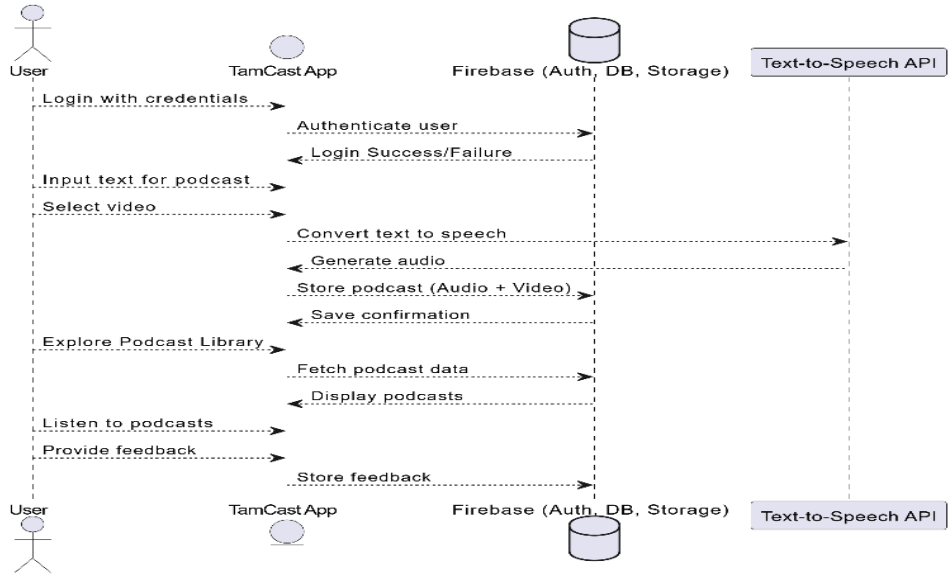

Fig 1: Sequence Diagram

The podcast storage and retrieval system ensures that all user-generated podcasts are organized within the "My Podcast" section, providing easy access, retrieval, and management. Firebase Cloud Storage supports real-time media storage and playback without latency issues. To further enhance engagement, TamCast integrates a user feedback system, allowing users to rate and review podcasts, provide content suggestions, and enhance recommendations, thereby improving overall content quality. Testing and quality assurance play a crucial role in refining the application. Unit testing validates individual modules such as TTS conversion, video integration, and Firebase storage, while integration testing ensures seamless interaction between different components. Additionally, user testing is conducted to gather feedback and refine the interface before deployment.

Once testing is complete, TamCast is deployed on the Google Play Store, where Firebase Analytics is used to track performance and gather user insights. Continuous updates, bug fixes, and feature enhancements are implemented based on real-time user feedback. By following this structured methodology, TamCast effectively combines AI-driven narration, video integration, and cloud-based podcast storage to deliver an intuitive, scalable, and innovative podcasting solution. Its continuous refinement cycle ensures seamless functionality, improved user engagement, and long-term sustainability.

IV. FUTURE SCOPE

The TamCast aims to evolve into a fully AI-driven podcasting platform, expanding beyond basic Text-to-Speech (TTS) conversion. Future enhancements include multi-language support, voice cloning, and AI-powered audio refinements for more natural narration. Content summarization could provide quick podcast highlights, while ensures copyright protection. Monetization can be expanded with premium hosting, sponsorships, and ad placements. Additionally, social media sharing and live podcasting features would boost engagement, making TamCast a versatile, secure, and profitable solution for content creators worldwide.

V. CONCLUSION

TamCast simplifies podcast creation with AI-powered Text-to-Speech (TTS), video integration, and secure cloud storage, ensuring efficiency and accessibility. Its WebView-based Podcast Library enhances content discovery, while Firebase supports seamless authentication and data management. Looking ahead, TamCast can expand with multi-language support, voice cloning, AI-driven enhancements, and more for better personalization and security.

Features like content summarization, live podcasting, and social media sharing will boost engagement, while monetization options create revenue opportunities. With its scalable and AI-driven approach, TamCast is set to redefine digital podcasting, making it more accessible, innovative, and engaging for creators worldwide.

REFERENCES

[1]. H. Zen, A. Senior, and M. Schuster, "Statistical Parametric Speech Synthesis Using Deep Neural Networks," IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 2, pp. 297–309, Apr. 2014, doi:10.1109/JSTSP.2014.2305763.

[2]. J. Shen, R. Pang, R. J. Weiss, M. Schuster, and N. Jaitly, "Natural TTS Synthesis by Conditioning Wavenet on Mel Spectrogram Predictions," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 4779–4783, doi:10.1109/ICASSP.2018.8461368.

[3]. C. Olston, J. Najork, M. R. Henzinger, and A. Paepcke, "Web Crawling," Foundations and Trends in Information Retrieval, vol. 4, no. 3, pp. 175–246, 2010, doi:10.1561/1500000017.

[4]. J. Dean and L. A. Barroso, "The Tail at Scale," Communications of the ACM, vol. 56, no. 2, pp. 74–80, Feb. 2013, doi:10.1145/2408776.2408794.

[5]. N. Eyal, "Hooked: How to Build Habit-Forming Products," IEEE Consumer Electronics Magazine, vol. 3, no. 2, pp. 21–23, Apr. 2014, doi:10.1109/MCE.2014.2304051.

[6]. A. van den Oord, S. Dieleman, H. Zen, and K. Simonyan, "WaveNet: A Generative Model for Raw Audio," IEEE Transactions on Audio, Speech, and Language Processing, vol. 25, no. 12, pp. 2292–2305, Dec. 2017, doi:10.1109/TASLP.2017.2734096.

[7]. A. Graves, N. Jaitly, and A. Mohamed, "Hybrid Speech Recognition with Deep Bidirectional LSTM," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, pp. 273–277, doi:10.1109/ICASSP.2013.6639345.

[8]. P. Taylor, "Text-to-Speech Synthesis," IEEE Signal Processing Magazine, vol. 25, no. 3, pp. 123–125, 2008, doi:10.1109/MSP.2008.918416.

[9]. H. Hermansky and N. Morgan, "RASTA Processing of Speech," IEEE Transactions on Speech and Audio Processing, vol. 2, no. 4, pp. 578–589, Oct. 1994, doi:10.1109/89.326616.

[10]. A. Karpathy, G. Toderici, S. Shetty, and T. Leung, "Large-Scale Video Classification with Convolutional Neural Networks," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1725–1732, doi:10.1109/CVPR.2014.223.

[11]. X. Huang, J. Baker, and R. Reddy, "A Historical Perspective of Speech Recognition," Communications of the ACM, vol. 57, no. 1, pp. 94–103, Jan. 2014, doi:10.1145/2500887.

[12]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, and L. Kaiser, "Attention is All You Need," Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[13]. P. Ladefoged and K. Johnson, "A Course in Phonetics," IEEE Transactions on Education, vol. 45, no. 3, pp. 157–160, Aug. 2002, doi:10.1109/TE.2002.1024592.

[14]. G. E. Hinton, S. Osindero, and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Nets," Neural Computation, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi:10.1162/neco.2006.18.7.1527.

[15]. K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770–778, doi:10.1109/CVPR.2016.90.

[16]. Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A Neural Probabilistic Language Model," Journal of Machine Learning Research, vol. 3, pp. 1137–1155, 2003.

[17]. D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," International Conference on Learning Representations (ICLR), 2015.

[18]. H. Sak, A. Senior, and F. Beaufays, "Long Short-Term Memory Recurrent Neural Network Architectures for Large Scale Acoustic Modeling," Annual Conference of the International Speech Communication Association (INTERSPEECH), 2014.

[19]. A. Graves, A. Mohamed, and G. Hinton, "Speech Recognition with Deep Recurrent Neural Networks," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2013, pp. 6645–6649, doi:10.1109/ICASSP.2013.6638947.

[20]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788, doi:10.1109/CVPR.2016.91.