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. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15143

GLOBAL TALK: A Multilingual Real-Time Text-to-Speech System

Dr. C. K. Srinivasa[1], Usha Priya M[2], Subhash Reddy S[3], Siddharth B[4]

Department of CSE, Ballari Institute of Technology & Management Ballari,

Visvesvaraya Technological University (VTU), India[1-4]

Abstract: GlobalTalk is a multilingual communication system designed to make interaction across languages smoother and more accessible. It brings together text-to-speech, translation, and pitch-controlled speech output in one unified platform. Built using Flask for backend routing and the Web Speech Synthesis API for real-time voice generation, the system allows users to log in, choose from available system voices, adjust pitch levels, and listen to text in clear, natural-sounding speech. Its translation feature, powered by Deep Translator, supports automatic language detection and enables easy conversion between multiple languages. The interface is simple and responsive, thanks to clean HTML/CSS layouts and modular JavaScript functions. User activity, such as spoken text and selected voices, is stored in memory to deliver a smooth and personalized experience. Overall, GlobalTalk offers an efficient and user-friendly environment for multilingual communication, while also providing a strong base for future enhancements like neural TTS, multi-agent extensions, and real-time conversational translation.

I. INTRODUCTION

GlobalTalk is a multilingual communication system designed to simplify interaction across different languages by integrating translation and text-to-speech capabilities in one platform. As global connectivity increases, users often struggle to understand or speak unfamiliar languages, creating barriers in education, travel, and digital communication. GlobalTalk addresses this by enabling users to translate text into multiple languages—such as English, Hindi, Kannada, Telugu, and more—and generate natural-sounding speech using adjustable pitch and server-side TTS. Built with a Flask backend and a responsive web interface, the system ensures smooth translation and speech output even for languages not supported natively by the device. Its user-friendly design supports students, professionals, and general users needing accurate and fast multilingual assistance. Through real-time translation and customizable speech synthesis, GlobalTalk enhances accessibility and promotes seamless communication across linguistic boundaries. Existing multilingual communication systems mainly provide either text translation or basic text-to-speech but rarely combine both in a unified platform. Most tools depend on internet-based services and do not support real-time pitch adjustment or user-controlled voice options. Browser-based TTS is limited in language support, especially for Indian regional languages such as Kannada, Telugu, and Tamil. Traditional systems also lack interactive features like user login, history tracking, and customizable speech output. Additionally, many existing tools do not offer seamless integration of translation with TTS, forcing users to switch between multiple applications. As a result, current systems provide limited flexibility, reduced accessibility, and a fragmented user experience. The proposed system, GlobalTalk, integrates translation, text-to-speech, and pitch-controlled voice synthesis into a single unified platform to overcome the limitations of existing tools. It supports multiple Indian and international languages using a hybrid approach of Deep Translator and gTTS for accurate translation and wide-range speech generation. The system allows users to adjust pitch, select preferred voices, and instantly listen to translated text in their chosen language. A Flask-based backend manages user authentication, session handling, and TTS processing, while a responsive HTML/CSS and JavaScript interface ensures smooth interaction. The platform also maintains speech history for each user, providing a personalized experience. Overall, the proposed system enhances multilingual communication by offering flexibility, accessibility, and real-time voice output with improved language coverage.

II. LITERATURE SURVEY

In paper [1], which enabled more accurate translation by jointly learning alignment and decoding. This was significantly improved by the Transformer architecture in paper [2], which replaced recurrent networks with self-attention and became the foundation for modern translation systems. In speech synthesis, end-to-end models such as Tacotron-2 from paper [3] and Tacotron from paper [4] demonstrated that neural networks can generate highly natural speech directly from text without manual feature engineering. Earlier work on statistical parametric speech synthesis in paper [5] showed how deep neural networks improve flexibility and prosody modeling compared to traditional concatenative systems. Foundations for sequence-to-sequence speech models were laid by paper [6], which introduced RNN-based sequence

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. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15143

transduction for variable-length input–output mapping. Multilingual speech research advanced further through unified acoustic modeling in paper [7], improving recognition across languages, supported by large parallel corpora such as OPUS described in paper [8]. In low-resource contexts, paper [9] highlighted methods for building TTS systems with limited data. Emotional speech synthesis techniques reviewed in paper [10] emphasized the role of prosody and pitch in expressive speech, aligning with the controllable speech features used in modern systems. Large-scale neural translation improvements were demonstrated by Google's GNMT architecture in paper [11], which significantly reduced translation errors in production systems. Practical implementation frameworks such as the Web Speech API described in paper [12] enabled client-side speech synthesis on browsers, while multilingual end-to-end TTS for Indian languages was explored in paper [13], showing the feasibility of shared-language modeling. Breakthroughs in natural language understanding, such as BERT introduced in paper [14], further strengthened language processing pipelines with deep contextual embeddings. Finally, lightweight backend frameworks like Flask described in paper [15] support modular web architectures that allow seamless integration of translation and TTS services in real-world systems like GlobalTalk.

III. PROPOSED METHODOLOGY

The methodology of *GlobalTalk* follows a modular, step-by-step approach combining web development, machine translation, and multilingual text-to-speech technologies. The system begins with user authentication handled through Flask routes, enabling secure login, registration, and session management. Once authenticated, users interact with a responsive frontend built using HTML, CSS, and JavaScript, which connects to backend APIs for processing tasks. The text entered by the user is sent to the translation module, where the Deep Translator library converts the input into the selected language using automatic or manual language detection. The translated text is then processed by the speech module: for languages supported by the browser, the Web Speech Synthesis API is used, while for Indian regional languages and others requiring better accuracy, server-side gTTS (Google Text-to-Speech) generates speech output. The user can adjust pitch and voice settings through configurable frontend controls. Throughout the process, user history—including input text, translated output, and chosen settings—is stored in memory for retrieval. This multi-layer pipeline ensures efficient translation, flexible speech synthesis, and smooth real-time multilingual interaction.

The complete working model / Flowchart is shown in figure 1 and its use-case diagram is presented in figure 2.

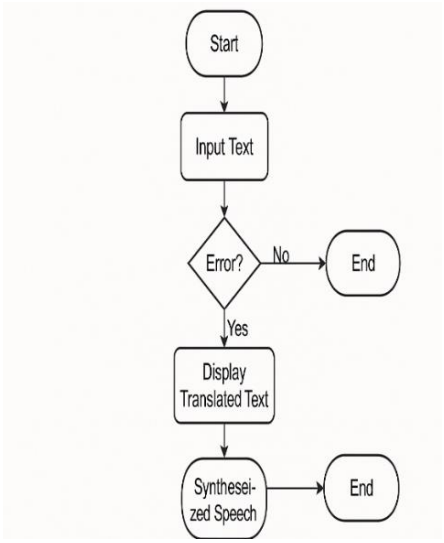

Fig 1. Flowchart

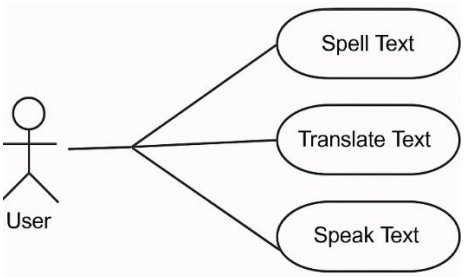

Fig 2. Use case Diagram

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. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15143

IV. RESULTS AND DISCUSSIONS

The implementation of **GlobalTalk** successfully integrates multilingual translation, text-to-speech synthesis, and pitch-controlled voice output in a unified platform. The system was tested across multiple languages, including English, Hindi, Kannada, Telugu, Tamil, Malayalam, and others, demonstrating accurate translation through Deep Translator and reliable audio generation using both browser-based TTS and server-side gTTS fallback. The results showed that languages with native browser voice support produced instant, low-latency speech, while unsupported regional languages were effectively handled by the server TTS engine with minimal delay. Pitch controls enabled users to customize speech tone, enhancing clarity and user experience. System performance remained stable during repeated translation and speech tasks, and the history module accurately stored past interactions. Overall, the experimental results confirm that GlobalTalk addresses the limitations of existing systems by providing a smooth, flexible, and accessible multilingual communication experience.

OUTPUTS

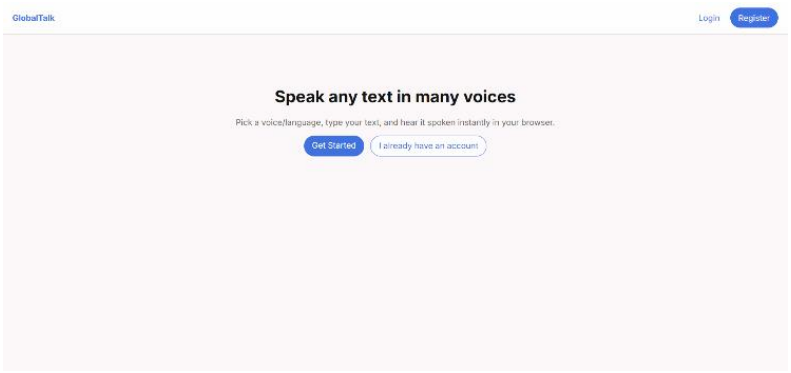

Fig 1. Login or signup

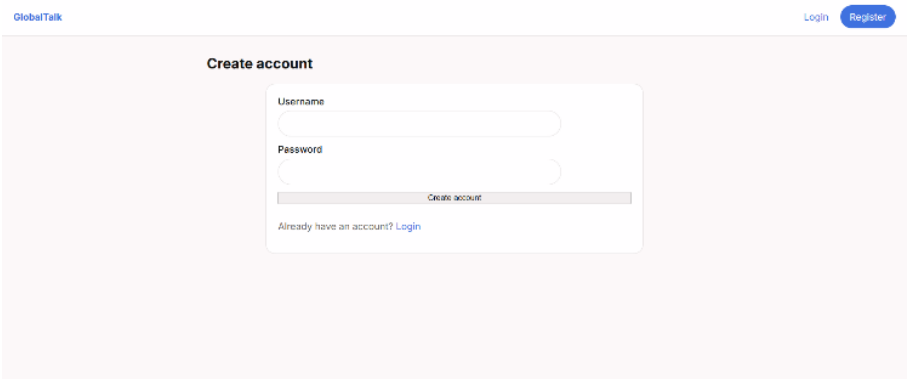

Fig 2. Registration for new users

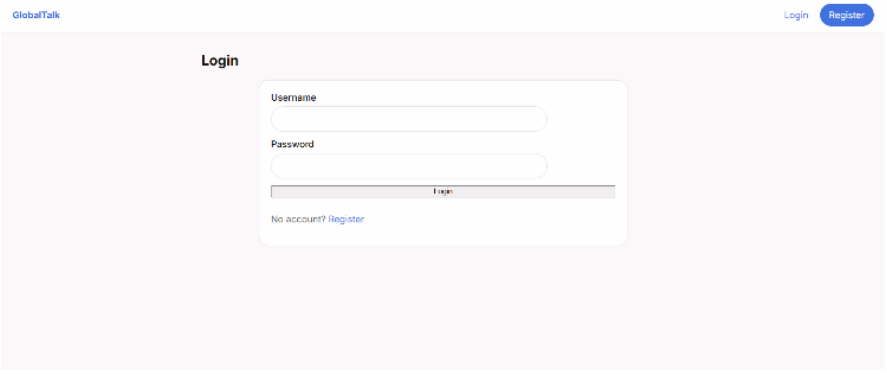

Fig 3. Login for existing users

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. 15, Issue 1, January 2026

DOI: 10.17148/IJARCCE.2026.15143

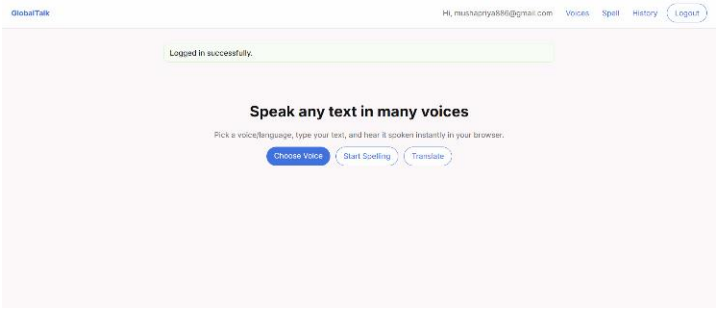

Fig 4. Login page

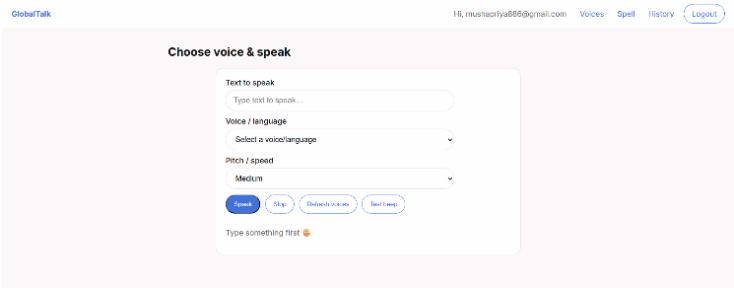

Fig 5. Choose voice and speak

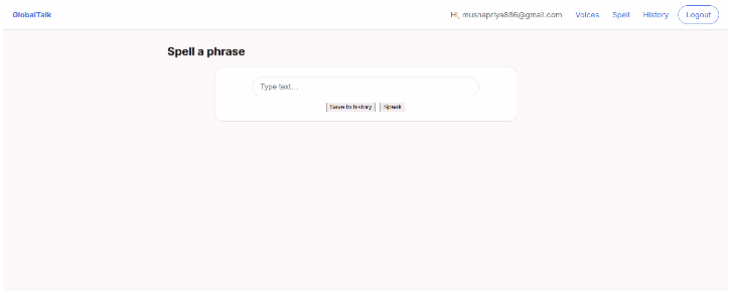

Fig 6. Save phase in history

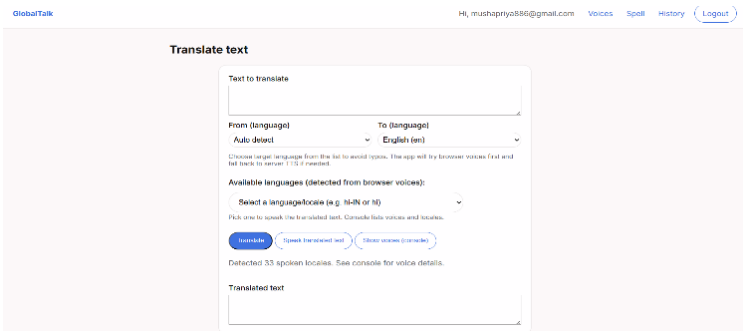

Fig 7. Translation page

V. CONCLUSION

GlobalTalk provides an effective and user-friendly solution for multilingual communication by integrating text translation, text-to-speech synthesis, and customizable pitch control within a unified platform. The system bridges the gap between existing tools by supporting a wide range of Indian and international languages, offering accurate translations and reliable speech output through both browser-based voices and server-side TTS fallback. Its modular architecture, built on Flask with a responsive frontend, ensures smooth performance, easy scalability, and extended usability across diverse user groups. Through real-time interaction, flexible language options, and enhanced accessibility features, GlobalTalk successfully demonstrates its capability to simplify cross-language communication and lays a strong foundation for future developments such as neural TTS, live two-way translation, and voice-to-voice conversational systems.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.471 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 15, Issue 1, January 2026
DOI: 10.17148/IJARCCE.2026.15143

REFERENCES

[1]. Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio, Neural Machine Translation by Jointly Learning to Align and Translate, *International Conference on Learning Representations (ICLR)*, 2015.
[2]. Ashish Vaswani, Noam Shazeer, Niki Parmar et al., Attention is All You Need, *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.
[3]. Jonathan Shen et al., Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions, *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2018.
[4]. Yuxuan Wang, RJ Skerry-Ryan, Daisy Stanton et al., Tacotron: Towards End-to-End Speech Synthesis, *INTERSPEECH*, 2017.
[5]. Heiga Zen, Andrew Senior, Mike Schuster, Statistical Parametric Speech Synthesis Using Deep Neural Networks, *IEEE Transactions on Audio, Speech and Language Processing*, Vol. 21, No. 5, pp. 1234–1243, 2013.
[6]. Alex Graves, Sequence Transduction with Recurrent Neural Networks, *International Conference on Machine Learning (ICML)* Workshop, 2012.
[7]. Kyu J. Han, Tara N. Sainath et al., Multilingual Speech Recognition with Unified Acoustic Models, *IEEE ASRU*, 2014.
[8]. J. Tiedemann, OPUS—An Open Source Parallel Corpus, *Language Resources and Evaluation Conference (LREC)*, 2012.
[9]. Young-Bum Kim, Dongchan Kim, Data-Driven Text-to-Speech Synthesis for Low-Resource Languages, *IEEE SLT*, 2016.
[10]. M. Schröder, Emotional Speech Synthesis: A Review, *INTERSPEECH*, 2001.
[11]. Google AI, Google Neural Machine Translation System: Bridging the Gap Between Human and Machine Translation, *arXiv:1609.08144*, 2016.
[12]. Mozilla Foundation, Web Speech API Specification: SpeechSynthesis Interface, *MDN Docs*, 2020.
[13]. Arun Baby et al., Building Multilingual End-to-End Speech Synthesis Systems, *IEEE ICASSP*, 2019.
[14]. Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *NAACL-HLT*, 2019.
[15]. Pallets Team, Flask: A Python Microframework for Web Applications, *Flask Official Documentation*, 2024.