IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Animated Multi-Lingual Voice & Text Bot For Seamless Interaction

Mrudula S R [1], Adithi R [2], Arvind N [3], G C Sambram [4]and Lakshmi K K [5]

Student, Department of AI&ML, K S Institute of Technology, Bengaluru, Karnataka, India [1-4]

Assistant Professor, Department of AI&ML, K S Institute of Technology, Bengaluru, Karnataka, India [5]

Abstract: With the advancement of artificial intelligence (AI) and natural language processing (NLP), chatbots have evolved into essential tools for multilingual and multi-modal communication. This paper presents an animated multilingual voice and text bot that integrates real-time language translation and speech synthesis for seamless human-computer interaction. The proposed system leverages neural machine translation (NMT) and deep learning-based text-to-speech (TTS) synthesis, ensuring accurate, real-time conversational experiences. The inclusion of animated facial expressions enhances user engagement, particularly for diverse linguistic users. This research explores the architecture, methodology, and implementation of the bot and discusses experimental results demonstrating its effectiveness in bridging language barriers.

Index Terms - Multilingual Chatbot, Neural Machine Translation, Text-to-Speech, Animated Conversational Agents, Natural Language Processing.

I. INTRODUCTION

In today's interconnected world, effective multilingual communication is essential across businesses, education, healthcare, and customer support. While traditional text-based chatbots have helped bridge language gaps, they often lack natural, engaging, and expressive interactions. To address these limitations, we propose an animated multilingual chatbot that supports both text and voice communication with enhanced realism and context-awareness.

Our system integrates neural machine translation (NMT) models and animated avatars to deliver accurate, real-time translations and lifelike interactions. Animated avatars featuring facial expressions, lip synchronization, and gestures provide vital non-verbal cues, improving comprehension and engagement.

Key objectives include:

1. **Real-Time Multilingual Communication** through text and voice.
2. **Animated Avatars** for human-like interactions with visual expressions.
3. **Neural Machine Translation (NMT)** for accurate, context-aware translations.
4. **Deep Learning-Based Text-to-Speech (TTS)** for natural-sounding multilingual voice outputs.
5. **User Interaction Studies** to evaluate translation accuracy, engagement, and satisfaction.

By combining advanced translation, TTS models, and expressive avatars, the chatbot aims to create a more immersive, accessible, and effective communication experience across linguistic and cultural boundaries.

II. RELATED WORK

Several studies have explored the development of multilingual chatbots and real-time translation systems, highlighting advancements in natural language processing (NLP), machine translation, and conversational AI. These studies form the foundation for our research, which aims to enhance chatbot interactions by integrating animation and real-time voice synthesis for a more immersive user experience. This section reviews key contributions and shows how our approach builds upon these advancements.

[1] Developing a Secure and Multilingual Chat App with Real-Time Translation

This study explores chatbot implementations using BERT-based models for Indian languages, presenting a framework for efficient multilingual conversation handling. It leverages BERT to improve language understanding, translation accuracy, and sentiment analysis through fine-tuning on Indian datasets.

A key contribution is its hybrid approach combining rule-based translation with machine learning for complex structures. Our project aligns with this by using advanced NLP but differentiates itself by integrating animated avatars and voice synthesis, enhancing the chatbot's dynamism and user-friendliness through visual and auditory communication cues.

[2] Enhancing Natural Language Processing in Multilingual Chatbots for Cross-Cultural Communication

This research highlights the role of deep learning and transformer-based models in enhancing chatbot communication across languages and cultures. It emphasizes the use of contextual embeddings and attention mechanisms to refine chatbot understanding and generation of human-like responses.

Our work extends these advancements by adding animated avatars that display facial expressions and gestures, providing visual context to multilingual conversations. By combining NMT models with visual enhancements, we aim to create a more intuitive and culturally sensitive communication experience.

[10] Multilingual Chatbot for Indian Languages

This study addresses the challenges posed by Indian languages' complex grammar and script variations. It showcases how real-time translation can bridge communication gaps and discusses the importance of encryption and secure authentication to protect sensitive chat data.

Inspired by this research, our project also emphasizes data security in real-time translation. However, we further enrich user engagement by introducing animated avatars and voice synthesis, improving accessibility for users with limited literacy and offering a more natural, inclusive interaction experience.

Our Approach and Contribution

Building on these studies, our work introduces a novel approach by integrating animation and real-time voice synthesis into multilingual chatbots. While prior research focused mainly on NLP, translation accuracy, and security, we enhance user engagement with expressive avatars and synchronized speech.

This integration of NLU with visual and auditory elements makes conversations more intuitive, engaging, and accessible, especially for users with visual impairments or low literacy. Our approach aims to set a new benchmark for conversational AI in cross-lingual communication.

III. METHODOLOGY

The methodology outlines the architecture and key modules that enable real-time multilingual communication through animated, voice-assisted interactions. It details the system’s core components, multilingual processing capabilities, and avatar integration. Together, these elements create an immersive and accessible user experience.

3.1 System Architecture

The proposed system comprises six core modules that collaboratively enable real-time, multilingual voice and text interaction. The architecture is designed to process spoken and written input, translate language accurately, and render responses using animated avatars for enhanced user engagement.

Speech Recognition (ASR)

The system utilizes Automatic Speech Recognition (ASR) to convert spoken language into text. ASR is powered by advanced deep learning models, such as end-to-end Transformer-based architectures and hybrid Deep Neural Network-Hidden Markov Models (DNN-HMM). These models enhance speech recognition accuracy by leveraging phonetic features, contextual understanding, and noise filtering.

Natural Language Understanding (NLU)

The Natural Language Understanding (NLU) module interprets this textual input to extract intent and relevant entities. It uses multilingual pre-trained models such as MuRIL BERT and XLM-R to ensure consistent performance across languages.

Neural Machine Translation (NMT)

The Neural Machine Translation (NMT) module is responsible for real-time translation of recognized text. It employs transformer-based models capable of dynamic language detection and context-aware translation, preserving both semantic accuracy and cultural relevance. Pre-trained models like multilingual BART and MuRIL BERT enhance the fidelity of translated output.

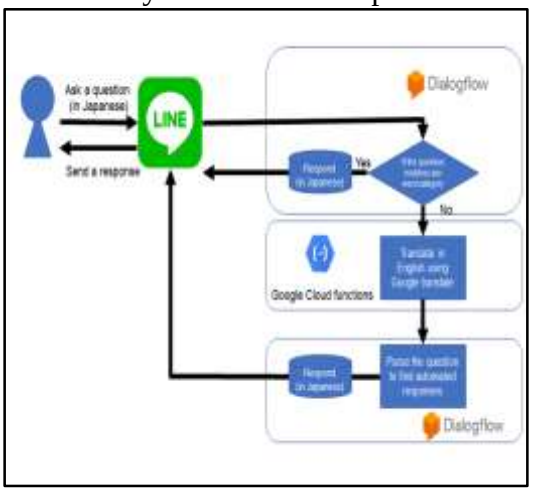

Figure III.1 interconnected components of system architecture of the animated multilingual voice and text bot

Text-to-Speech (TTS) Synthesis

Text-to-speech (TTS) synthesis converts translated text into spoken output. The TTS system supports a wide range of languages and modulates voice characteristics, tone, and prosody to produce natural-sounding speech. It can also adapt to regional dialects and user preferences to improve communication accessibility.

Animation Module

To create an immersive experience, an AI-driven animation module synchronizes speech with realistic facial expressions and gestures. Lip-sync is achieved through phoneme alignment, while facial expressions and movements are dynamically generated based on sentiment analysis. This module enhances the naturalness and relatability of the interaction.

Dialog Manager

At the core, the dialog manager integrates inputs and outputs from all modules to maintain coherent, multi-turn conversations. It manages context retention, handles user-specific preferences, and ensures a seamless conversational flow by adapting responses based on the evolving state of the interaction.

3.2 Multilingual Processing

The system is designed for robust multilingual communication, ensuring inclusivity and accessibility. It dynamically detects the user's spoken or written language without requiring manual selection. By leveraging contextual information from surrounding dialogue, the system preserves grammatical accuracy and semantic meaning during translation.

Personalization plays a critical role in multilingual processing. The system learns from user interactions, adapting translations to align with regional dialects, speech patterns, and historical preferences. It also supports code-switching, enabling users to naturally shift between languages within a single conversation, a common feature in many multilingual communities.

3.3 Animated Avatar Integration

The animated avatar plays a central role in creating a more engaging and human-like interface. It uses deep learning-based facial animation models to reflect speech and emotion in real time. Lip-sync is handled through phoneme-to-viseme mapping, enabling accurate and language-agnostic mouth movements.

In addition to lip-sync, the avatar supports non-verbal communication cues such as head movements, hand gestures, and gaze control, all of which contribute to a more lifelike presence. These gestures are driven by the conversational context and the emotional tone of the user's input. Users can also customize the avatar's appearance and expressions, increasing personalization and comfort during interaction.

IV. IMPLEMENTATION

The system was implemented using a combination of state-of-the-art technologies to provide seamless real-time interaction. The key components involved in the development include programming languages, speech and translation APIs, avatar animation tools, and cloud services, ensuring a highly responsive and efficient chatbot system. The chatbot supports both text and voice inputs, making it accessible across multiple platforms, including web and mobile applications.

4.1 Programming Languages and Frameworks

Python served as the primary backend language, facilitating natural language processing and integration of deep learning models. TensorFlow and PyTorch were used for developing and experimenting with the Neural Machine Translation (NMT) models. The frontend was developed using React.js, enabling a dynamic and responsive user interface.It enables seamless integration with APIs and real-time updates, making interactions more engaging and responsive.

4.2 Speech Recognition and Translation APIs

Speech recognition was handled by OpenAI's Whisper model, offering robust multilingual transcription. Google Text-to-Speech (TTS) generated natural voice responses, while DeepL API was used for context-aware, high-precision translations.

Table IV.1 working flow chart of the proposed model

Input Collection	• Speech-to-Text (STT) Tools • Microphone Integration (Hardware Support)
Language Detection	• Language Detection APIs (e.g., Google Cloud or LangDetect)
Translation to Target Language	• Machine Translation APIs (e.g., Google Translate, AWS Translate)
Natural Language Processing (NLP)	• NLP Libraries (e.g., SpaCy, NLTK, or Transformer-based models like GPT)
Output Generation - Text	• UI Frameworks for Text Display (e.g., Tkinter, React, Flask)
Output Generation - Voice	• Text-to-Speech (TTS) Tools (e.g., gTTS, Google TTS, or pyttsx3)
Feedback and Iterative Improvement	• User Feedback Mechanism (Logging, API for Responses) • Model Fine-Tuning Frameworks (e.g., Hugging Face)

4.3 Avatar Animation and Real-Time Rendering

Unity 3D powered real-time rendering of animated avatars, creating expressive and lifelike interactions. Daz3D was employed to design detailed 3D character models, integrated seamlessly into Unity for real-time animation and lip-syncing.

4.4 Cloud Services and Database Management

AWS provided scalable backend infrastructure, with services like AWS Lambda and S3 managing logic and assets. Firebase was utilized for real-time database management, supporting synchronized, multi-platform user experiences.

4.5 User Interaction and Platform Accessibility

The chatbot supports text and voice-based inputs, offering cross-platform compatibility through web and mobile applications. Users can interact naturally, receiving real-time translations and responses, thus bridging linguistic barriers effectively.

The combination of these technologies enables a highly efficient, user-friendly, and interactive chatbot system, bridging language barriers and enhancing communication across different linguistic

backgrounds.The chatbot interface supports both **text-based and voice-based inputs**, allowing users to interact through a **web application** and **mobile platforms**.

V. Results and Discussion

The chatbot was evaluated through a structured user study involving 30 participants fluent in Hindi, Kannada, and English. Participants completed five tasks including real-time translations, voice-based queries, and avatar-guided interactions. Metrics evaluated included translation accuracy (BLEU score), speech recognition (Word Error Rate, WER), system response time, and user engagement (Likert-scale surveys).

5.1 Quantitative Results

A one-way ANOVA showed a significant difference in engagement levels across systems ($F(2,87) = 6.78$, $p < 0.01$). Post-hoc Tukey tests indicated that our system outperformed Google and DeepL-based setups ($p < 0.05$). BLEU score achieved was 0.81, WER was 9.2%, and average engagement rating was 4.5 out of 5, demonstrating high translation and recognition accuracy.

Table V.1 quantitative results

Metric	Our System	Google Translate + TTS	DeepL + pyttsx3
BLEU Score	0.81	0.76	0.83
WER	9.2%	13.6%	10.5%
Response Time (s)	1.2	1.0	1.4
Engagement Rating	4.5	3.2	4.1

5.2 User Study Observations

87% of participants preferred the animated avatar. Visual feedback improved comprehension, particularly for non-native speakers. Some users (13%) noted lip-sync mismatches under network lag, and 27% suggested enhancing emotional tone in speech synthesis.

5.3 Summary and Limitations

The system achieved robust performance but requires improvement in emotional TTS, dialect handling, and avatar synchronization. Future work will focus on scaling user studies and enhancing emotional responsiveness.

VI. Conclusions and Future Work

This study presents an innovative approach to multilingual conversational AI by integrating real-time language translation, speech synthesis, and animation. The developed chatbot enhances user interaction by providing a dynamic, engaging, and immersive experience that goes beyond conventional text-based solutions. By leveraging Natural Language Processing (NLP) and Neural Machine Translation (NMT), the chatbot ensures seamless communication across multiple languages, breaking language barriers and fostering inclusivity. Additionally, the incorporation of text-to-speech (TTS) and animated avatars further enriches user engagement, making interactions more lifelike and natural.

While the current implementation demonstrates significant advancements in multilingual AI-driven conversation, there are several potential areas for future improvement and expansion:

- **Emotion Recognition AI**: Future iterations of the chatbot can incorporate emotion recognition technology to enhance its responsiveness. By analyzing vocal tone, facial expressions, and textual sentiment, the chatbot could adapt its responses in a more empathetic and context-aware manner, improving user satisfaction and fostering deeper connections.
- **Expanded Language Support**: The addition of more regional languages and dialects would make the chatbot more versatile and globally applicable. By continuously updating its language models and integrating lesser-known dialects, the chatbot can cater to a more diverse user base, ensuring that language barriers do not hinder effective communication.

• **Cloud Optimization**: To improve performance and scalability, future enhancements should focus on cloud-based deployment and optimization. Implementing efficient cloud computing solutions would reduce processing delays, ensure seamless real-time translation, and support a higher number of concurrent users. Optimizing server loads and using distributed AI processing techniques would enhance the chatbot's responsiveness and availability.

In summary, this work lays a strong foundation for an advanced multilingual AI chatbot that enhances communication and engagement through real-time translation and animated interaction. By incorporating emotion recognition, gesture-based interactions, expanded linguistic support, and cloud optimization, future developments will further refine and elevate the chatbot's capabilities, making it a more powerful tool for seamless global communication.

REFERENCES

[1] Pandey, P., Ashtankar, S., Datir, S., Agre, O., Wazalwar, S., Thombre, S. S. (2024). Developing a Secure and Multilingual Chat App with Real-Time Translation. Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS).

[2] Orosoo, M., Goswami, I., Alphonse, F. R., Fatma, G., Rengarajan, M., Bala, B. K. (2024). Enhancing Natural Language Processing in Multilingual Chatbots for Cross-Cultural Communication. 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV).

[3] Benita, J., Kumar, R. M. R., Elambharati, E., Reddy, G. M., Raj, K. N. (2024). Multi-Language Conversational Agent for Tech Support: Design and Implementation. International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA).

[4] Karanam, S. R., Balaji, A. S. N., Swaroop, V. V., Sai, V. S., Manda, R., Chowdary, P. A. (2024). M-Bot: Bridging the Gap in Mine Safety for Non-Literate Workers through Multilingual Chatbots Assistance. International Conference on Computational Intelligence for Green and Sustainable Technologies (ICCIGST).

[5] Sinthusha, A. V. A., Charles, E. Y. A., Weerasinghe, R. (2024). Machine Reading Comprehension for the Tamil Language With Translated SQuAD. IEEE Access.

[6] Radhika, A., Bhasin, N. K., Sabareesh, R., Raju, Y. R., Satyanarayana, K. N. V., Raj, I. I. (2024). Radak, Optimization of Natural Language Processing Models for Multilingual Legal Document Analysis. Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS).

[7] Deshmukh, P., Kulkarni, N., Kulkarni, S., Manghani, K., Khadkikar, P. A., Joshi, R. (2024). Named Entity Recognition for Indic Languages: A Comprehensive Survey. 1st International Conference on Trends in Engineering Systems and Technologies (ICTEST).

[8] Reedy, M. V., Naeem, A. B., Reddy, T. K., Vinusha, B. V., Pramila, R. P. (2024). Natural Language Translation Engine for Announcements and Information Dissemination at Stations. Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE).

[9] Shirisha, N., Srivani, M., Kowsalyadevi, K., Ram, G. B. P., Vadrevu, P. K., Murthy, V. B. (2024). Real-Time Multilingual Farming Assistance using NLP Integrated Web API. 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS).

[10] Singh, U., Vora, N., Lohia, P., Sharma, Y., Bhatia, A., Tiwari, K. (2023). Multilingual Chatbot for Indian Languages. 14th International Conference on Computing Communication and Networking Technologies (ICCCNT).