

Archives available at [journals.mriindia.com](http://journals.mriindia.com)

## International Journal on Advanced Electrical and Computer Engineering

ISSN: 2349-9338

Volume 14 Issue 01, 2025

### AI Powered Voice Agent using NLP and ASR

<sup>1</sup>Prof. S. O. Bagnawar, <sup>2</sup>A. S. Rananavare, <sup>3</sup>R. V. Rananavare, <sup>4</sup>D. D. Tarange, <sup>5</sup>R. D. Thombare

<sup>1 2 3 4 5</sup>Department of AI & DS, SBPCOE, Indapur, India

Email: <sup>1</sup>swatiwaghmode111@gmail.com, <sup>2</sup>argroup060@gmail.com, <sup>3</sup>rohanrananavare0811@gmail.com, <sup>4</sup>tarangednyaneshwari10@gmail.com, <sup>5</sup>rushikeshthombare286@gmail.com

Peer Review Information	Abstract
<p><i>Submission: 11 Sept 2025</i></p> <p><i>Revision: 10 Oct 2025</i></p> <p><i>Acceptance: 22 Oct 2025</i></p> <p><b>Keywords</b></p> <p><i>Voice Agent, Natural Language Processing (NLP), Automatic Speech Recognition (ASR), Call Center Automation, Multilingual AI, Large Language Models (LLMs), N8N.</i></p>	<p>The demand for intelligent, automated communication systems has significantly increased with the growth of businesses and customer service requirements. Traditional call centers require high manpower, leading to increased operational costs and inconsistent service quality. This paper surveys the development of AI-powered voice agents using Natural Language Processing (NLP) and Automatic Speech Recognition (ASR) for automating call center operations. The system supports Marathi, Hindi, and English, enabling multilingual communication and bridging the gap between technology and regional language users.</p> <p>A key feature of this work is the integration of the <b>N8N automation platform</b>, which enables modular workflows for call handling, appointment booking, and marketing processes. By leveraging N8N, the system can seamlessly connect ASR, NLP, and Large Language Models (LLMs) such as ChatGPT and Gemini, ensuring flexibility, scalability, and ease of deployment. This modular approach reduces the complexity of development while allowing customization for businesses of different sizes.</p> <p>We discuss the background, related works, methodologies, performance results from existing studies, and future research directions. The survey highlights how voice-based AI agents, integrated with N8N and modern LLMs, can transform customer support, marketing, and appointment management by reducing human dependency and operational costs.</p>

### Introduction

Customer support remains one of the most resource-intensive functions for businesses worldwide. With the rapid growth of digital services, customers expect quick, accurate, and personalized responses. Traditional human-operated call centers often face challenges such as long waiting times, limited availability, and inconsistent service quality. The rise of artificial intelligence (AI) has enabled the development of voice-enabled agents capable of handling customer queries,

booking appointments, and providing multilingual support around the clock.

Unlike conventional interactive voice response (IVR) systems, which rely on pre-programmed menus, AI-powered agents utilize NLP and ASR to interpret natural speech, respond intelligently, and adapt to various accents and dialects. This paper provides a survey of AI-based voice agents, focusing on multilingual capabilities (Marathi, Hindi, English) and integration with modern Large Language Models (LLMs). Additionally, we

emphasize the role of automation platforms like N8N in simplifying workflow integration and enabling businesses of different scales to adopt AI-powered solutions.

### Background

Artificial intelligence-based voice agents draw from several foundational technologies:

- **Automatic Speech Recognition (ASR):** Converts spoken input into text, forming the core layer for further natural language processing.
- **Natural Language Processing (NLP):** Interprets the meaning of the transcribed speech by detecting intent and context.
- **Large Language Models (LLMs):** Advanced AI models such as ChatGPT and Gemini enhance dialogue quality by generating context-aware and human-like responses.
- **Multilingual Challenges:** Indian languages pose specific challenges due to phonetic diversity, regional accents, and scarcity of annotated datasets for training models.
- **Automation Platforms:** Tools like N8N provide modular integration for AI workflows, enabling seamless appointment booking, marketing campaigns, and CRM connectivity.

### Related Work

Several industrial and academic solutions have been explored in the domain of conversational AI and call center automation:

- Google Dialogflow and Amazon Lex have successfully demonstrated conversational agents capable of automating customer support at scale.
- Studies on multilingual ASR stress the necessity of accent adaptation, particularly for Indian subcontinent languages.
- Integration of LLMs with dialogue systems has shown considerable improvements in coherence and naturalness of responses compared to rule-based and retrieval-based approaches.
- Research on customer service automation highlights reductions in operational cost and enhanced customer satisfaction due to improved response quality and availability.

### Literature Survey

Several notable works have contributed to the field:

1. Sarikaya et al. (2017) proposed conversational AI architectures designed for task-oriented dialogue systems, providing a foundation for practical deployment in customer-facing environments.
2. Kumar et al. (2020) presented multilingual ASR systems tailored for Indian languages, employing deep learning techniques to handle phonetic diversity.
3. Budzianowski et al. (2018) introduced the MultiWOZ dataset, which has become a benchmark for task-oriented dialogue modeling, enabling advancements in context management.
4. More recent studies (2021–2024) explore hybrid approaches combining ASR, NLP, and LLMs, demonstrating superior performance for multilingual and multi-domain call center applications.

### Methodology

The proposed AI-powered voice agent system is structured with the following pipeline:

- **Speech-to-Text Conversion:** Input speech is processed through ASR models trained on multilingual datasets.
- **Language Understanding:** NLP modules extract user intent, context, and relevant entities from the transcribed text.
- **LLM Integration:** Models such as ChatGPT and Gemini generate context-aware responses, adapting to business-specific needs.
- **Text-to-Speech (TTS):** Synthesizes natural-sounding audio responses in the chosen language.
- **N8N Workflow Automation:** Orchestrates background processes such as appointment booking, call routing, and logging interactions.
- **Multilingual and Accent Adaptation:** Ensures effective performance across Marathi, Hindi, and English, incorporating fine-tuned models for Indian speech patterns.

### Performance Result

Performance evaluation data from prior works and prototypes is summarized in Table 1.

Table 1: Performance Metrics of AI-Powered Voice Agent

Metric	Hindi	English	Marathi
ASR Accuracy (Word Level)	90%	88%	78%
Avg. Response Latency	1.5s	1.4s	1.6s
Operational Cost Reduction	Up to 60%		
Customer Satisfaction Increase	Over 40%		
Concurrent Calls Supported	Thousands		

### Future Directions

Although existing advancements are promising, several challenges remain:

- Expanding annotated datasets for underrepresented regional languages, particularly Marathi.
- Enhancing models to adapt dynamically to rural, urban, and mixed-accent speech patterns.
- Incorporating offline capabilities with lightweight deployment models for areas with limited connectivity.
- Integrating emotion and sentiment recognition to allow empathetic, human-like interactions.
- Seamless integration with CRM, marketing tools, and other enterprise systems.
- Broadening language support to include additional Indian and global languages to make the system universally adaptable.

### References

S. Zhang, Y. Liu, and K. Chen, "Multilingual Non-Autoregressive ASR with LLM Fusion," in *Proc. IEEE ICASSP*, 2024, pp. 4657–4661. [IEEE Xplore]

M. Wu, J. Shi, and C. Li, "ARoBERT: A Pre-Trained Language Model Robust to ASR Errors," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 1120–1132, 2022. [IEEE Xplore]

H. Wang, X. Li, and Y. Qian, "Integrating Multiple ASR Systems: Attention-Based Fusion of Multiple ASR Outputs for Robust NLP," in *Proc. IEEE ICASSP*, 2022, pp. 5672–5676. [IEEE Xplore]

R. Sharma and P. Kumar, "NLP-Based Sanskrit

Voice Bot Using ASR and NLP," in *Proc. IEEE Int. Conf. Smart Tech.*, 2021, pp. 311–316. [IEEE Xplore]

M. Brown, L. Rossi, and E. Garcia, "SELMA: Speech-Enabled Large Language Models for Multi-Task Virtual Assistants," *arXiv preprint*, arXiv:2501.05678, 2025. [arXiv:2501.05678]

V. Ethiraj, A. David, S. Menon, and D. Vijay, "Toward Low-Latency End-to-End Voice Agents for Telecommunications Using Streaming ASR, Quantized LLMs, and Real-Time TTS," *arXiv preprint*, arXiv:2508.04721, 2025. [arXiv:2508.04721]

J. Park, A. Lee, and H. Kim, "TokenVerse: A Unified Model for ASR and NLP Tasks in a Single Transducer-Based Framework," *arXiv preprint*, arXiv:2403.01921, 2024. [arXiv:2403.01921]