IJAAR PUBLISHING®

Scholarly Journal of Science and Technology Research & Development | ISSN: 2955-0807
Vol. 5, Issue 2 (2026) | www.ijaar.org/sjstrd

A Context Aware Igbo Language Voice Assistant Using Natural Language Processing Tools

Emmanuel Nwabueze Ekwonwune[1], Leticia E. Elebiri[1],
Abraham Oghenemega Ovwonuri[2], Chibundu D. Asiegbu[2], Netochukwu Nwokafor[2]
[1]Department of Computer Science, Imo State University, Owerri
[2]Department of Software Engineering, Federal University of Technology, Owerri

Abstract

The aim of this research is to develop a prototypecontext aware voice assistant system for the Igbo language using natural language processing tools. This will enable human–computer interaction through spoken Igbo commands. The system combines Natural Language Processing (NLP) techniques and tools such as Automatic Speech Recognition (ASR), and Text-to-Speech (TTS) technologies to process, understand, and respond to users voice inputs/commands. Specifically, it uses NCAIR1/Igbo-ASR model for speech-to-text conversion, Gemini AI for translation and response generation, and YarnGPT for natural-sounding voice output. This research adopts a Component-Based Software Development Life Cycle (CBSDLC) by focusing on building the system from independently developed and reusable software components. A structured user evaluation was conducted with ten (10) native Igbo speakers in other to represent varying fluency levels and exposure to the system. Across all users, the system achieved an average ASR accuracy of 89.2%, corresponding to an average WER of 10.8%.Further analysis of intent recognition performance yielded an average precision of 0.87, recall of 0.85, and F1 score of 0.86. The assistant achieved a Task Success Rate (TSR) of 82%, indicating that most user intents were successfully completed without breakdown. Similarly, the Dialogue Completion Rate reached 80%, confirming that the majority of interactions concluded with satisfactory responses.The assistant recorded an average response time of 2.81 seconds. The Response Appropriateness Score, measured on a five-point Likert scale, achieved an average rating of 4.1, indicating that responses were generally relevant, culturally appropriate, and linguistically coherent.Additionally, the system achieved an average Mean Opinion Score of 3.8, reflecting good overall speech intelligibility and naturalness.

Keywords: Context aware, Voice Assistant, Automatic Speech Recognition (ASR), Text-to-Speech (TTS), Natural Language Processing (NLP), Igbo Language

https://doi.org/10.46654/hby7p316

1.0 Introduction

Voice assistant system is a system that enables users to interact with their computer using voice inputs/commands. Voice assistants have become an important aspect in modern human-computer interaction because it enables users to perform tasks using human understandable voice commands. These systems are possible due to advancements in Automatic Speech Recognition (ASR), Natural Language Processing (NLP), and Artificial Intelligence (AI). They are embedded in devices and applications that facilitate realtime, no-touch, hands-free interaction with digital platforms (Amour & Devi, 2025). These sophisticated programs have fully integrated widely spoken international languages such as English, Spanish, and French. However, lowresource languages like Igbo have not yet been fully integrated hence the drive for this research.

There are multiple examples of voice assistant systems such as Siri, Tesla voice control, ChatGPT with voice (OpenAI), and so on. Each of these voice assistants serves a unique purpose; therefore, they are not interchangeable. However, they are not fully optimized for Igbo language processing and often produce inaccurate pronunciation of Igbo words due to the limitations of their underlying speech models. At times, they also misinterpret input by mistakenly processing Igbo statements as if they were another language.

The Igbo language is a tonal language belonging to the kwa group of the Niger-Congo language family (Fulford, 2002). It is spoken by the Igbo people, primarily residing in southeastern Nigeria but also present in other parts of the country and in various diaspora communities (Oluikpe, 2014). The Igbo language has more than 30 dialects and this variance is so much that certain dialects of this same Igbo language are unintelligible to each other, hence the standard Igbo (Igbo Izugbe) that serves as a central dialect for Igbo people (Clark, 1990). This research adopts Igbo Izugbe (Macgregor, 1909)exclusively, as its standardized nature is expected to enhance accessibility and attract wider interest across diverse Igbo-speaking communities.

Currently, there is a massive global adoption for voice assistant in various languages and especially tonal languages (Shazhaev et al., 2023). This study builds upon current available resources to design a voice assistant system that integrates Igbo Izugbe- the central Igbo language that is the general official Igbo language for teaching in schools and writing Igbo books - to facilitate Igbo language inclusion in the digital sphere.

1.1 Related Works

Several studies have been carried on voice assistant systems, we have therefore grouped them into seven (7) subheadings based on architectural design, application context and functionality.

1.1.1 Mobile and Smartphone-Based Voice Assistants

These studies focus on voice assistants designed for mobile platforms. They emphasizing hands-free interaction, accessibility, and real-world usability. Studies, such asKiwa et al. (2023) on AI Voice Assistant for Smartphones with NLP Techniques, Raghavan et al. (2021) on Virtual AI Assistant for Person with Partial Vision Impairment and Noorullah et al. (2025) on AI-Powered Multi-Lingual Voice Interactive Cooking Assistant, fall into this category. Their main focus is on mobility, accessibility, hands free interaction, real-time assistance. They demonstrated how voice assistants can enhance accessibility and hands-free interaction in mobile systems, particularly for visually impaired users. They also demonstrated how voice assistants can be beneficial in task-focused scenarios. While these systems show strong

real-time responsiveness, most of them rely on cloud services and are focused on dominant languages. This exposes limitations for low-connectivity regions and under-represented languages like Igbo language.

1.1.2 Desktop and System Control Voice Assistants

This category covers assistants designed for desktop environments which emphasizes automation, productivity, and control at system level. Such works byGupta et al. (2022) on Python-based "Personal voice assistant", Tiwari & Patel (2024) on an online AI based voice assistant using Python and PyTorch, Kannan et al. (2024) on Jarvis for Windows with GPT integration,Aggarwal et al. (2024)on Desktop voice assistant, Mukadam et al. (2025)on Jarvis a Voice Assistant for Desktop and Masoom Basha (2025) on Voice-Enabled Personal Desktop Assistant for System Control and Automation fall into this category. Their basic focus is on productivity, automation, personalization, system control. These systems have demonstrated the feasibility of Python based architectures for managing local applications and workflows but most of them are optimized for English and depend on pre-trained models with limited cultural adaptability. Hence exposes their limitation for low resource languages like Igbo language.

1.1.3 Multilingual and Low Resource Language Voice Assistants

These works directly addressed multilingual support, language dataset scarcity, and adaptation to low resource linguistic contexts. Works by Algondi. A (2025) on Multilingual assistant for rural healthcare, Noorullah et al. (2025) on Multilingual cooking assistant,R. K. Singh et al. (2025) on Enhancing Voice Assistant Systems through Advanced AI and NLP Techniques, Lazzaroni et al. (2024) on embedded end-to-end voice assistant and Held et al. (2024) on Distilling an end-to-end voice assistant without instruction training data. They focused on low resource languages, dataset scarcity, robustness in multilingual scenerio. They try to address the challenges of supporting multiple languages, especially those with limited annotated datasets. These studies emphasize adaptive models, targeted data collection, and transfer learning to maintain performance. Studies like this lay a solid foundation for the voice assistant systems in low resource languages like Igbo language.

1.1.4 LLM-Integrated and ContextAware Voice Assistants

These studies build upon the power of large language models (LLM) for enhanced conversational intelligence and contextual understanding. Studies by H. Singh et al. (2024) on Bing GPT assistant with LSTM and attention, Mukherjee et al. (2025) on LLM-based industrial voice dialogue system, Amour & Devi (2025) on An intelligent model for a virtual Voice Assistant using modern artificial intelligence technologies and Held et al. (2024) on Distilling an End-t-End voice assistant without instruction training data are under this category. Their key focus was on conversational depth, contextual reasoning, LLM integration. LLM powered assistant systems have demonstrated enhanced conversational depth, contextual reasoning, and task flexibility. However, they are often computationally expensive and depends on cloud thereby limiting accessibility in low resource environments.

1.1.5 Accessibility Focused Voice Assistants

These studies are focused on voice assistants that are designed to support users with disabilities through perception and voice feedback. Such as the studies by Kiwa et al. (2023) on AI Voice Assistant for Smartphones with NLP Techniques, Raghavan et al. (2021) on Virtual AI Assistant for Person with Partial Vision Impairment, Natuskar et al. (2025) on Voice Assistant for the Blind, Austerjost et al. (2018)on Introducing a Virtual Assistant to the Lab and Hombeck et al. (2024)on voice user interface (VUI) for effortless navigation in

medical virtual reality environment are this category. Their primary focus was on inclusivity, safety, perception-based assistance. Their works have illustrated how voice interfaces can support users with visual, physical, or situational impairments through speech based feedback and hands free control. These assistants emphasis the social value of inclusive design.

1.1.6 Multimodal and Perception Enhanced Voice Assistants

These studies focused on systems that combine speech with vision or other modalities to improve context awareness. Works by Vinoth kumar R et al. (2025) on Voice based virtual assistant using python, Raghavan et al. (2021) on Virtual AI Assistant for Person with Partial Vision Impairment, ,Austerjost et al. (2018)on Introducing a Virtual Assistant to the Lab and Hombeck et al. (2024) on voice user interface (VUI) for effortless navigation in medical virtual reality environment belong to this category. Their primary focus is on multimodal interaction, environmental awareness. Multimodal assistant systems that integrate speech with vision or environmental perception have demonstrated to have improved contextual awareness. Though effective, such systems increases complexity and computational cost.

1.1.7 Review, Survey, and Conceptual Studies

These works provide foundational knowledge, ethical considerations, and system evolution insights. Works by Shazhaev et al. (2023) on Personal Voice Assistant from Inception to Everyday Application and Austerjost et al. (2018) on Introducing a Virtual Assistant to the Lab: A Voice User Interface for the Intuitive Control of Laboratory Instruments fall on this category. They focused on system evolution, ethics, comparative analysis. Review based studies provide critical insights into the evolution, ethical challenges, and architectural patterns of voice assistants. They highlighted persistent issues such as privacy, continuous listening, and language bias.

1.2 Methodology

This project adopts a Component-Based Software Development Life Cycle (CBSDLC). The following phases (figure 1) outline the CBSDLC as applied in this project.

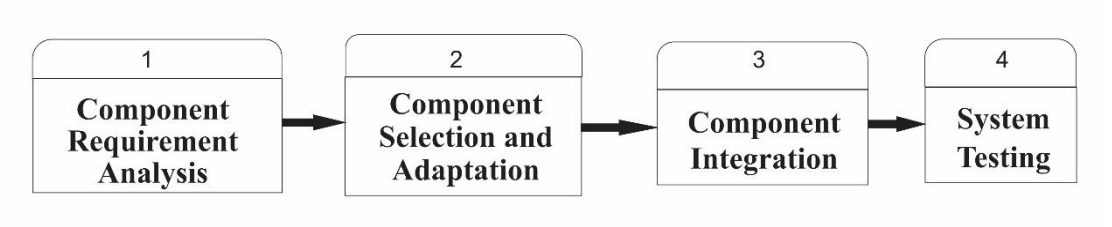


Figure 1: Methodological Flow Diagram

Phase 1 - Component Requirement Analysis: The system was broken into distinct functional modules (e.g., ASR engine, TTS engine, user interface, and data logger). Each component was clearly defined by its responsibilities and communication with other parts of the system.

Phase 2 - Component Selection and Adaptation: Pre-trained open-source components such NCAIR1/Igbo-ASR for ASR and YarnGPT TTS for speech synthesis were selected. Tools like Python, Hugging Face Transformers.

Phase 3 - Component Integration: Selected components were integrated to work together seamlessly. The ASR module converts Igbo speech to text, which is processed and optionally converted back to speech using a TTS module.

Phase 4 - System Testing: Components were tested individually (unit testing) and together (integration testing). The entire system was evaluated at the end of each sprint, and improvements were made based on feedback.

1.2.1 System Design/ Architecture

The system is organized into four primary modules/components (Figure 2), each encapsulating a major function of the voice assistant system.

1. Speech Input Module**:** Captures real-time Igbo speech through a microphone and forwards the audio to the ASR engine.
2. ASR Component (Speech-to-Text)**:** Utilizes a NCAIR1/Igbo-ASR model to transcribe Igbo speech into text.
3. ResponseComponent: Utilizes Gemini AI model to translate and understand the input before generating a response in Igbo.
4. TTS Component (Text-to-Speech)**:** Uses YarnGPT model to convert text responses to Igbo speech.

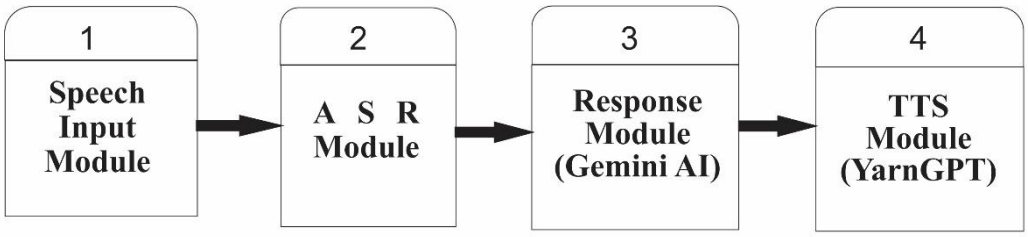


Figure 2: Architecture of the Igbo Voice Assistant

Figure 2 shows how the different modules interact to achieve intelligent voice communication in the Igbo language. The design allows each module (ASR, Translation and Reasoning, and Text-to-Speech) to function as an independent component while working together through an orchestration layer. This architecture emphasizes modularity, flexibility, and real-time performance. Each component can be replaced or updated without disrupting the entire system

1.3 Results and Discussion

Following the development of the Igbo Voice Assistant, a structured user evaluation was conducted with ten (10) native Igbo speakers in other to represent varying fluency levels and exposure to the system. Each participant interacted with the system using spontaneous Igbo utterances, which included greetings, weather inquiries, and general conversational prompts. The evaluation adopted a multi-layered performance framework, covering speech recognition accuracy, dialogue efficiency, task completion, and perceived response quality.

1.3.1 Automatic Speech Recognition (ASR) Performance Evaluation

The ASR component was evaluated using Word Error Rate (WER) together with recognition accuracy or Word Accuracy (WAcc) as they provide a more detailed assessment of transcription performance.

The Word Error Rate (WER) is a crucial metric used to evaluate the performance and accuracy of aASR system as it quantifies the number of mistakes the TTS makes when translating or transcribing a sequence of words compared to a correct transcript. A lower WER indicates a better ASR accuracy. It is calculated by comparing the system-generated output to a correct transcript and counting three types of errors.

WER = (S + D + I) / N

Where:

S = Substitutions: defines the number of words that where incorrectly replaced by another word.

D = Deletions: define the number of words that were in the correct transcript but missed or omitted by the system.

I = Insertions: defines the number of extra words added by the system that were not the correct transcript.

N = Total Words: Total number of words in the original correct transcript (N= S + D + C, where C is the number of correct words).

The Word Accuracy (WAcc) measures how often the system correctly transcribes or translates spoken or written words. It is the complement of WER.

WAcc = 1 – WER. Expressed in percentage as:

WAcc (%) = (1 – WER) x 100%

Its most common formular is:

WAcc = (C – I) / N

Where:

C = defined as the number of correct words (words correctly transcribe and aligned)

I = Insertions: defines the number of extra words added by the system that were not the correct transcript.

N = Total number of words in the original correct transcript.

Across all users, the system achieved an average ASR accuracy of 89.2%, corresponding to an average WER of 10.8%. This indicates that approximately one out of every ten words was incorrectly recognised, a reasonable outcome for a tonal, low-resource language such as Igbo.

Further analysis of intent recognition performance yielded an average precision of 0.87, recall of 0.85, and F1 score of 0.86, demonstrating balanced performance in correctly identifying user intents while minimising false recognitions. Minor reductions in recall were observed for longer utterances and tone-sensitive lexical items, suggesting that tonal variation remains a key challenge for ASR robustness. Figure 3 illustrates ASR accuracy and corresponding WER across participants, highlighting generally consistent recognition performance with slight variability attributable to pronunciation differences and speech rate.

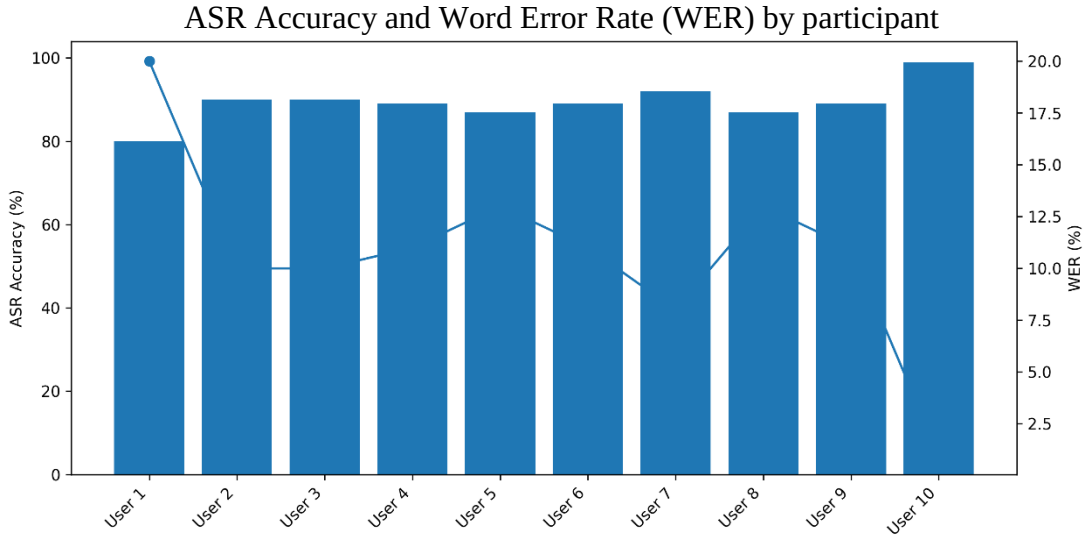


Figure 3: Automatic Speech Recognition (ASR) accuracy and corresponding Word Error Rate (WER) across participants

1.3.2 Task Performance and Dialogue Efficiency Evaluation

Beyond response accuracy, the system's ability to support goal-oriented interaction was assessed using dialogue-level metrics. The assistant achieved a Task Success Rate (TSR) of 82%, indicating that most user intents were successfully completed without breakdown. Similarly, the Dialogue Completion Rate reached 80%, confirming that the majority of interactions concluded with satisfactory responses.

The system required an average of 1.6 conversational turns per task, reflecting efficient interaction flow. However, instances of misunderstanding led to clarification requests, resulting in a Fallback / Clarification Rate of 18%. These fallback occurrences were primarily associated with tonal ambiguity and low-confidence ASR outputs, reinforcing the importance of tone-aware acoustic modelling in future system iterations.

Figure 4 presents task-level performance metrics, including TSR, dialogue completion rate, and fallback frequency.

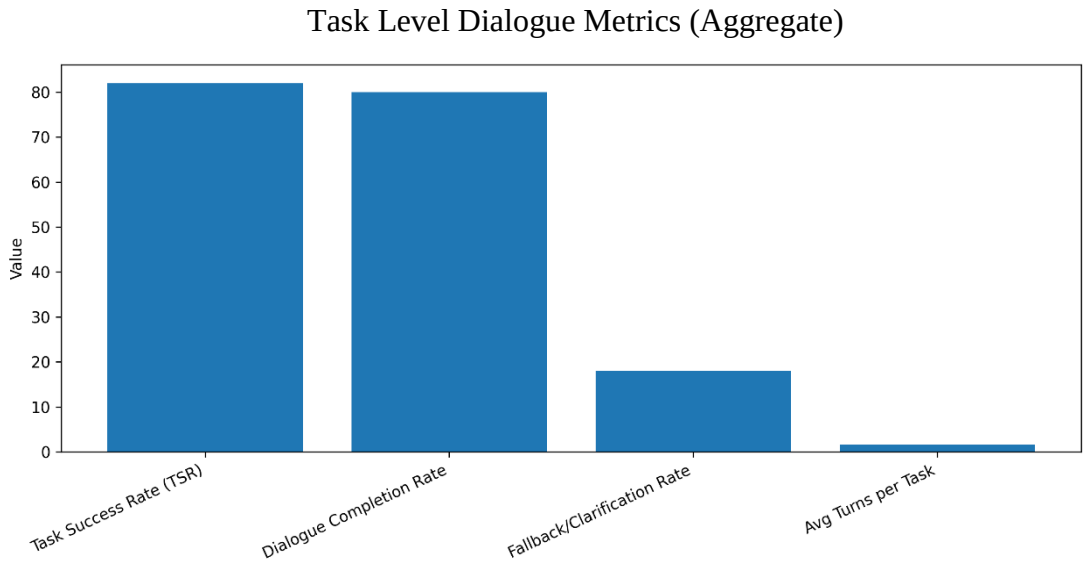


Figure 4: Task-level dialogue performance showing Task Success Rate, Dialogue Completion Rate, and Fallback Frequency.

1.3.3 Response Time and Interaction Responsiveness Evaluation

Response latency was evaluated to determine real-time interaction suitability. The assistant recorded an average response time of 2.81 seconds, as shown in Figure 5, which falls within acceptable thresholds for conversational voice systems. Users reported minimal disruption to conversational flow, suggesting that system processing and response generation were sufficiently optimised for interactive use.

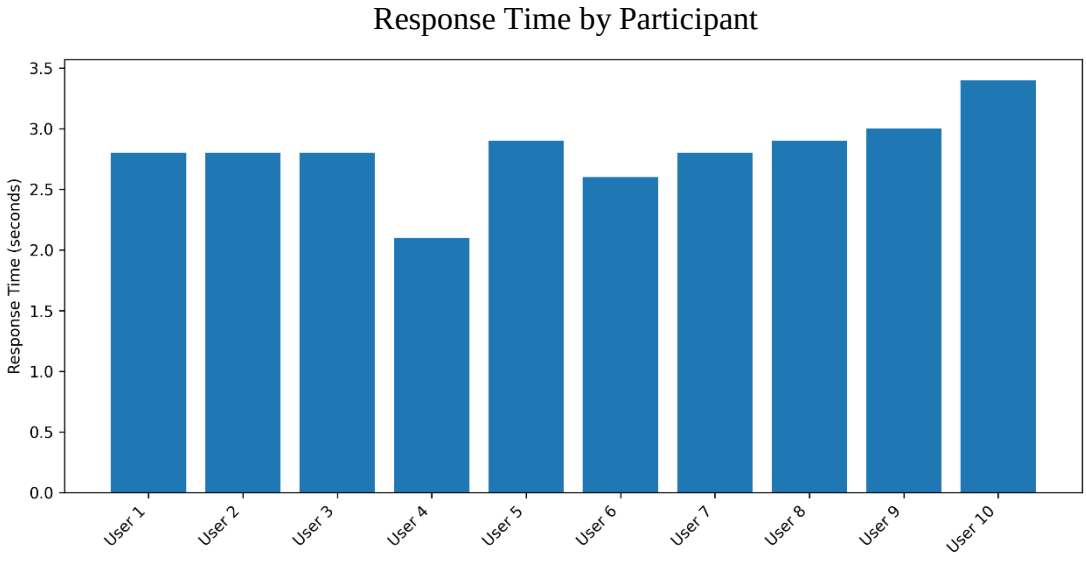


Figure 5: Average system response time per user during interactive voice sessions.

1.3.4 Response Quality and User Perception Evaluation

To assess response naturalness and appropriateness, participants rated system outputs using subjective evaluation metrics. The Response Appropriateness Score, measured on a five-point

Likert scale, achieved an average rating of 4.1, indicating that responses were generally relevant, culturally appropriate, and linguistically coherent.

Additionally, the Text-to-Speech (TTS) module was evaluated using Mean Opinion Score (MOS). The system achieved an average MOS of 3.8, reflecting good overall speech intelligibility and naturalness. Nonetheless, participants noted occasional tonal inconsistencies, particularly in sentence-final pitch contours and lexical tone realisation. These observations are consistent with known challenges in synthesising tonal languages under low-resource conditions. The mean opinion score and the response appropriateness score are illustrated in figure 6 below.

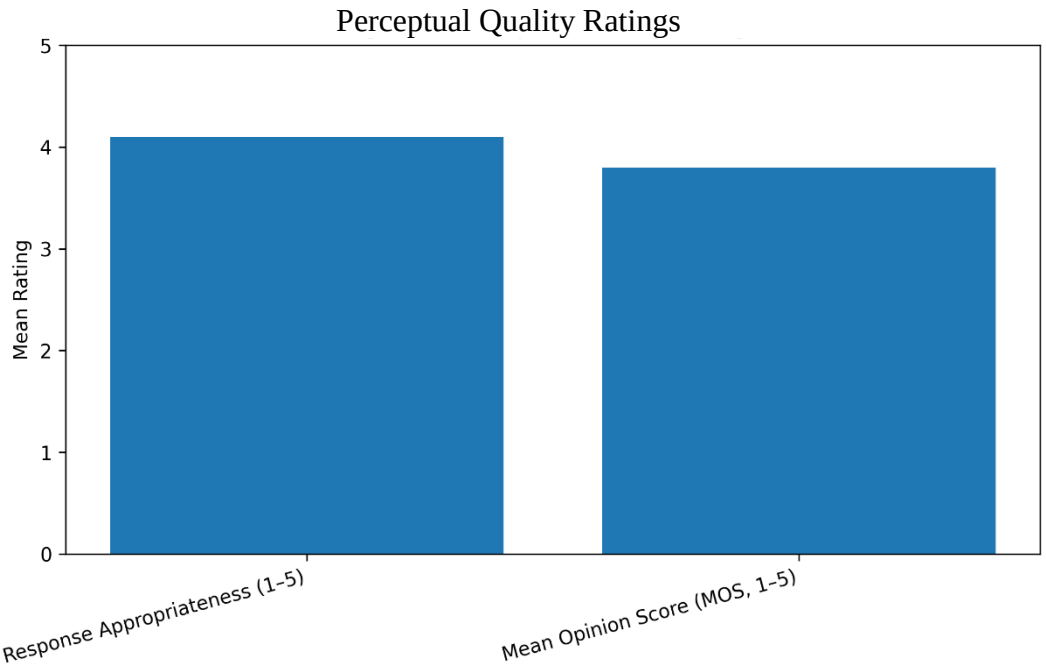


Figure 6: Perceptual Quality Ratings

Table 1: UserLevel ASR and Response Performance Metrics

User	ASR Accuracy (%)	Word Error Rate (%)	Response Time (Seconds)	Response Accuracy (%)
User 1	80	20	2.8	73
User 2	90	10	2.8	80
User 3	90	10	2.8	78
User 4	89	11	2.1	89
User 5	87	13	2.9	78
User 6	89	11	2.6	89
User 7	92	8	2.8	90
User 8	87	13	2.9	99
User 9	89	11	3	76
User 10	99	1	3.4	78
AVERAGE	**89.2**	**10.8**	**2.81**	**83**

Table 2: Aggregate Dialogue and Perceptual Evaluation Metrics

Metric	**Value**
Precision	0.87
Recall	0.85
F1 Score	0.86
Task Success Rate (TSR)	82%
Dialogue Completion Rate	80%
Average Turns per Task	1.6
Fallback / Clarification Rate	18%
Response Appropriateness Score (1–5)	4.1
Mean Opinion Score (MOS)	3.8

1.3.5 Overall Discussion

Collectively, the results demonstrate that the Igbo Voice Assistant achieves robust speech recognition, efficient dialogue management, and high user acceptance, despite inherent challenges associated with tonal modelling and limited training data. The integration of task-level and perceptual metrics provides a holistic evaluation, revealing that moderate ASR errors do not necessarily translate into task failure when supported by effective dialogue management strategies.

The findings confirm the system's potential as a functional and culturally relevant conversational agent, contributing to the advancement of inclusive voice technologies for under-resourced African languages.

1.4 **Conclusion**

The successful completion of this study proves that integrating low-resource African languages such as Igbo into intelligent systems is achievable using modern AI and component-based methodologies. The Igbo Language Voice Assistant System embodies the convergence of ASR, NLP, and TTS technologies within a unified architecture capable of recognizing, understanding, and responding to Igbo speech in real time.

Beyond its technical achievement, this study represents an important milestone in language preservation and digital inclusion. It addresses the linguistic gap in existing commercial assistants and promotes the participation of Igbo speakers in the digital economy through voice-driven interaction.

In conclusion, the study demonstrates that with open-source resources, pre-trained multilingual models, and iterative agile development, it is possible to design scalable and culturally relevant systems that give indigenous languages a meaningful presence in modern computing.

1.5 **References**

Aggarwal, M., Kumari, D., Kant, A., & Gupta, A. K. (2024). Desktop Voice Assistant. *International Journal for Research in Applied Science and Engineering Technology*, *12*(5), 1929–1934. https://doi.org/10.22214/ijraset.2024.61932

Algondi. A, N. (2025). Multilingual Voice AI Assistants for Bridging Language Gaps in Rural Healthcare. *International Journal of Scientific Research and Engineering Trends*, *11*(2), 2267–2271. https://doi.org/10.61137/ijsret.vol.11.issue2.421

Amour, A. S. AL, & Devi, S. G. (2025). An Integrated Model for A Virtual Voice Assistant Using Modern Artificial Intelligence Technologies. *INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)*, 1105–1112. https://doi.org/10.51583/IJLTEMAS

Austerjost, J., Porr, M., Riedel, N., Geier, D., Becker, T., Scheper, T., Marquard, D., Lindner, P., &Beutel, S. (2018). Introducing a Virtual Assistant to the Lab: A Voice User Interface for the Intuitive Control of Laboratory Instruments. *SLAS Technology*, *23*(5), 476–482. https://doi.org/10.1177/2472630318788040

Clark, M. M. (1990). The Tonal System of Igbo. In *The Tonal System of Igbo*. De Gruyter. https://doi.org/10.1515/9783110869095

Fulford, B. (2002). An Igbo Esperanto: A History of the Union Ibo Bible 1900-1950. *Journal of Religion in Africa*, *32*(4), 478. https://doi.org/10.1163/157006602321107658

Gupta, S., Joshi, D., Kolpate, K., Sharma, D., & Parmar, N. (2022). Personal Voice Assistant. *15th International Conference on Advances in Computing, Control, and Telecommunication Technologies, ACT 2024*, *10*(5), 62–67. https://doi.org/10.22214/ijraset.2022.42730

Held, W., Li, E., Ryan, M., Shi, W., Zhang, Y., & Yang, D. (2024). *Distilling an End-to-End Voice Assistant Without Instruction Training Data*. http://arxiv.org/abs/2410.02678

Hombeck, J., Voigt, H., & Lawonn, K. (2024). Voice user interfaces for effortless navigation in medical virtual reality environments. *Computers and Graphics (Pergamon)*, *124*. https://doi.org/10.1016/j.cag.2024.104069

Kannan, M. K. J., Jain, V., Kumar, A., Yadav, A. K., Verma, S., & Tiwari, R. (2024). Jarvis Voice Assistant. *International Journal of All Research Education and Scientific Methods (IJARESM)*, *12*.

Kiwa, F. J., Masamha, T., & Chinofunga, S. (2023, November). *AI voice assistant for smartphones with NLP techniques*.

Lazzaroni, L., Bellotti, F., & Berta, R. (2024). An embedded end-to-end voice assistant. *Engineering Applications of Artificial Intelligence*, *136*. https://doi.org/10.1016/j.engappai.2024.108998

Macgregor, J. (1909). Some Notes on Nsibidi. *Journal of the Royal Anthropological Institute*, *39*, 209–219. https://doi.org/10.2307/2843292

Masoom Basha, Mr. P. (2025). Voice-Enabled Personal Desktop Assistant for System Control and Automation. *INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT*, *09*(06), 1–9. https://doi.org/10.55041/ijsrem50673

Mukadam, Y., Baig, M. S., Seniwal, A., & Gharat, R. (2025). Jarvis – Voice Assistant for Desktop. *International Journal of Advanced Research in Science, Communication and Technology*, 78–79. https://doi.org/10.48175/ijarsct-23710

Mukherjee, A., Karande, A., Häfner, P., Poonia, M. D., Kimmig, A., Kreuzwieser, S., Vlas, R., Klar, M., Sykora, T., & Grethler, M. (2025). A LLM-based voice user interface for voice dialogues between user and industrial machines. *Procedia CIRP*, *134*, 378–383. https://doi.org/10.1016/j.procir.2025.02.138

Natuskar, S., Sable, A., Sharma, Y., Shende, S., & Itankar, P. Y. (2025). Voice Assistant for Blind. *International Journal for Research in Applied Science and Engineering Technology*, *13*(4), 3956–3960. https://doi.org/10.22214/ijraset.2025.69183

Noorullah, R. M., Anjum, S. H., Shivani, E., Begam, S. R., &Depuru, S. R. (2025). AI-Powered Multi-Lingual Voice Interactive Cooking Assistant. *International Journal of Innovative Research in Technology*, *12*(2), 636–644. https://www.researchgate.net/publication/394980518

Oluikpe, E. N. (2014). Igbo language research: Yesterday and today. *Language Matters*, *45*(1), 110–126. https://doi.org/10.1080/10228195.2013.860185

Raghavan, R., Krishnan, V., Nishad, H., & Shaikh, B. (2021). Virtual AI Assistant for Person with Partial Vision Impairment. *ITM Web of Conferences*, *37*, 01019. https://doi.org/10.1051/ITMCONF/20213701019

Shazhaev, I., Mikhaylov, D., Shafeeg, A., Tularov, A., & Shazhaev, I. (2023). Personal Voice Assistant: from Inception to Everyday Application. *Indonesian Journal of Data and Science*, *4*(2), 64–72. https://doi.org/10.56705/ijodas.v4i2.69

Singh, H., Kumari, D., Srinija, S., & Sarkar Bappy, S. B. (2024). Bing GPT Voice Assistant. *International Journal of Innovative Science and Research Technology (IJISRT)*, 1101–1107. https://doi.org/10.38124/ijisrt/ijisrt24aug373

Singh, R. K., Kathuria, S., Saraswat, P., & Kumar, A. (2025). Enhancing Voice Assistant Systems through Advanced AI and NLP Techniques. *Journal of Recent Innovations in Computer Science and Technology*, *2*(1), 52–62. https://doi.org/10.70454/jricst.2025.20110

Tiwari, P., & Patel, M. (2024). *Online AI Based Voice Assistant*.

Vinoth kumar R, Surendar R, Dhanush V, & Bharathkumar S. (2025). Voice Based Virtual Assistant Using Python. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, *3*(04), 1912–1916. https://doi.org/10.47392/irjaeh.2025.0278