METAVISION

Dr S Priyadharshini*, Dharani R, Akilesh J, Kamalesh V,Guruprasad S
*Assistant Professor, Department of Artificial Intelligence and Data Science, SNS College of Engineering(Autonomous), Coimbatore, India.
Research Scholars, Department of Artificial Intelligence and Data Science, SNS College of Engineering(Autonomous), Coimbatore, India.

***Abstract*—In the evolving hospitality industry, enhancing customer experience while maintaining operational efficiency has become a critical priority. Traditional food ordering systems in hotels and restaurants often depend on manual interaction, such as phone calls, waitstaff communication, or mobile text input, which can lead to delays, miscommunication, and order inaccuracies. To address these challenges, this project introduces MetaVisionn, an AI-powered voice-based smart ordering system.**

MetaVisionn enables customers to place food orders using natural speech through their mobile devices, eliminating the need for typing or manual input. The system utilizes advanced Speech-to-Text technology to convert spoken language into text in real time, followed by Natural Language Processing (NLP) techniques to extract key order details such as food items, quantities, special instructions, and delivery location (room or table number). The processed data is then structured and transmitted to hotel dashboards and kitchen display systems.

A unique aspect of MetaVisionn is its ability to present both the original spoken transcript and the AI-interpreted order, ensuring transparency and reducing errors in communication. The system also incorporates real-time streaming of spoken words, enhancing user interaction and engagement. By integrating AI-driven automation with intuitive voice interfaces, MetaVisionn significantly improves service speed, accuracy, and overall customer satisfaction, making it a scalable solution for modern smart hospitality environments.

I. INTRODUCTION

The hospitality sector has witnessed rapid technological advancements aimed at improving customer service and operational workflows. Despite these developments, many hotels and restaurants continue to rely on conventional ordering methods, including manual order-taking by staff or text-based mobile applications. These approaches often introduce inefficiencies such as longer waiting times, human errors, and communication gaps between customers, service staff, and kitchen teams.

With the increasing adoption of Artificial Intelligence (AI) and voice-enabled technologies, there is a growing opportunity to redefine how customers interact with service systems. Voice-based interfaces provide a natural, intuitive, and user-friendly method of communication, reducing dependency on manual input and minimizing friction in the ordering process. Customers can simply speak their requirements, making the interaction faster and more accessible.

In this context, MetaVisionn is proposed as an intelligent voice-driven ordering system specifically designed for hotel environments. The system captures user speech in real time, converts it into text, and applies Natural Language Processing (NLP) to understand the intent and extract structured order information. This information is then delivered seamlessly to hotel dashboards and kitchen displays.

Additionally, MetaVisionn enhances transparency by displaying both the raw spoken input and the processed output, allowing staff to verify and ensure accuracy. The system also incorporates features such as live speech visualization, waveform animation, and AI-based confirmation, creating a highly interactive and modern user experience.

By bridging the gap between human communication and digital processing, MetaVisionn aims to improve service efficiency, reduce operational errors, and provide a smarter, faster, and more reliable ordering experience in the hospitality industry.

II. RELATED WORKS

Several digital platforms and technologies have contributed to the evolution of food ordering and voice interaction systems. Online food delivery platforms such as Zomato and Swiggy have transformed how customers browse menus and place orders. These platforms offer features like digital

menus, real-time tracking, and online payments. However, they primarily rely on manual input methods such as typing or selecting items, which still require user effort and may not provide a fully seamless experience.

Voice-enabled technologies have also advanced significantly in recent years. Systems such as Amazon Alexa and Google Assistant demonstrate the effectiveness of speech recognition and natural language understanding in executing user commands. These systems allow users to perform tasks using voice interaction, such as setting reminders, controlling smart devices, and searching for information. However, their application in domain-specific environments like hotel ordering systems is limited and not optimized for extracting structured order data.

Research in speech recognition and NLP highlights the importance of real-time processing and contextual understanding for accurate interpretation of human language. Previous studies have explored chatbots and automated ordering systems, but most focus on text-based interfaces rather than continuous voice interaction.

MetaVisionn builds upon these existing technologies by integrating real-time speech-to-text conversion with domain-specific NLP tailored for hospitality services. Unlike traditional systems, it offers a fully voice-driven interface with live transcription, AI-based order extraction, and dual-layer visibility (spoken input and structured output). This approach addresses the limitations of current systems and provides a more efficient, intuitive, and error-resistant solution for modern hotel environments.

III. SYSTEM ARCHITECTURE

The proposed MetaVisionn system follows a modular and scalable architecture designed to efficiently process real-time voice inputs and deliver structured outputs to hotel management systems. The architecture is divided into three primary layers: Presentation Layer, Application Layer, and Data Layer.

Presentation Layer
The Presentation Layer acts as the user interface for different stakeholders, including guests, hotel staff, and kitchen personnel.

Guest Interface(Mobile/Web):
Provides a voice-first interaction system where users can place orders by speaking. It includes features such as a call button, real-time speech-to-text display, waveform animation, and dynamic visualization of spoken words.

Hotel Dashboard:
Displays incoming orders along with the original spoken transcript and AI-processed structured data. It allows staff to verify, manage, and process orders efficiently.

Kitchen Display System (KDS):Shows prepared order tickets along with the extracted order details and the original spoken input (“Guest said”). This ensures clarity and reduces misinterpretation during food preparation.

Application Layer
The Application Layer is the core processing unit of the system, responsible for handling all computational logic and AI-based operations.

Speech-to-Text Engine:
Converts real-time voice input into text using advanced speech recognition algorithms.

Natural Language Processing (NLP) Module:

Analyzes the converted text to identify key entities such as food items, quantities, special instructions, and delivery location.

Order Extraction Engine:
Structures the extracted information into a standardized format suitable for processing and display.

AI Response System:
Generates confirmation messages and interacts with the user if clarification is needed.

Notification System:
Sends real-time updates to the hotel dashboard and kitchen display.

Data Layer
The Data Layer is responsible for secure storage and management of all system data.
Stores:

User information
Voice transcripts
Processed orders
Feedback and logs
Uses cloud-based databases to ensure scalability, reliability, and data security.

AI Agent Workflow

The MetaVisionn system operates using an intelligent decision-making cycle:

Perception → Analysis → Decision Making → Action Execution → Learning

Perception: Captures voice input from the guest

Analysis: Converts speech to text and identifies intent

Decision Making: Extracts structured order details

Action Execution: Sends processed order to dashboard and kitchen

Learning: Improves system accuracy using feedback and usage data

This layered architecture ensures efficient real-time processing, scalability, and seamless integration between user interaction and backend intelligence.

IV. METHODOLOGIES

Requirement Analysis
The initial phase involves identifying the limitations of existing hotel ordering systems and gathering requirements for a voice-based solution.
Analyze current challenges such as manual errors and delays

Define functional requirements:
Voice input processing
Real-time transcription
Order extraction

Define non-functional requirements:
Low latency
High accuracy
Scalability and security

B. System Design
The system is designed with a voice-first approach, eliminating the need for traditional input methods.
UI/UX Design:
Focus on intuitive interaction with no typing interface
Architecture Design:
Three-layer architecture (Presentation, Application, Data)
Database Design:
Structured storage for orders, transcripts, and user data
NLP Model Design:
Entity recognition for extracting food items, quantities, and locations

C. Development Methodology
An Agile development approach is followed to allow iterative improvements and continuous feedback.
Frontend Development:

Built using HTML, CSS, JavaScript, or React for responsive UI
Backend Development:
Developed using Python (Flask/Django) or Node.js
AI Integration:
Implementation of speech recognition and NLP modules
API Integration:
Communication between frontend, backend, and AI services

D. Testing and Validation
Comprehensive testing is performed to ensure system performance and reliability.
Unit Testing:
Testing individual modules such as speech processing and NLP
Integration Testing:
Ensuring seamless communication between components
User Acceptance Testing (UAT):
Collecting feedback from real users
Performance Testing:
Evaluating system response time and accuracy under load

E. Deployment and Maintenance
The system is deployed on a cloud platform for scalability and accessibility.
Cloud hosting for real-time availability
Continuous monitoring of system performance
Regular updates and bug fixes
Feedback-driven improvements

F. Feedback and Iterative Improvement
Collect user feedback after each order
Analyze performance data
Improve AI accuracy and system efficiency over time

V. IMPLEMENTATION

The system is developed using HTML, CSS, and JavaScript for the frontend.

Backend is implemented using Python (Flask/Django) orNode.js.

The application captures user voice through a call/voice input feature.

Speech is converted into text using a Speech-to-Text engine.

NLP techniques are used to extract order details like items, quantity, and location.

Extracted data is structured and sent to the hotel dashboard and kitchen display.

A database (MySQL/Firebase) stores user data, orders, and transcripts.

APIs are used to connect frontend, backend, and AI modules.

The system is deployed on a cloud platform for real-time access.

VI. RESULTS AND ANALYSIS

The MetaVisionn system was tested to evaluate its performance in real-time hotel ordering scenarios. The results demonstrate significant improvements in efficiency, accuracy, and user experience compared to traditional ordering methods.

Performance Evaluation
Reduced Order Time:
Voice-based ordering significantly reduces the time required to place an order compared to manual typing or waiter-based systems.
Improved Accuracy:
The combination of speech recognition and NLP minimizes errors in order interpretation.

Real-Time Processing:
Orders are processed and displayed instantly on hotel dashboards and kitchen systems.

User Experience Analysis
Customers found the system easy to use and intuitive, as it mimics natural human conversation.
The absence of typing reduces effort and improves accessibility.
Real-time display of spoken words enhances user confidence and interaction

VII. ADVANTAGES

The MetaVisionn system offers several advantages in improving hotel food ordering processes. It provides a hands-free, voice-based ordering experience, eliminating the need for manual typing and making the system easy to use for customers. By using speech recognition and natural language processing, the system significantly reduces human errors and miscommunication between customers and staff. Real-time processing ensures that orders are delivered instantly to the hotel dashboard and kitchen display, resulting in faster service. Additionally, the system enhances transparency by displaying both the original spoken transcript and the AI-processed order, allowing staff to verify accuracy. Overall, MetaVisionn improves customer satisfaction, reduces workload on staff, and streamlines hotel operations.

VIII. FUTURE WORK

The MetaVisionn system can be further enhanced with several advanced features to improve its functionality and scalability. Future improvements include adding multi-language support to allow users to interact in regional languages such as Tamil and Hindi. Integration of voice-based payment systems can enable seamless transactions directly through speech. The system can also be upgraded with AI-based recommendation features to suggest food items based on user preferences. Integration with smart room and IoT devices can enable automated services within hotel environments. Additional enhancements such as noise cancellation and emotion detection can improve speech accuracy and prioritize urgent requests. The system can also be expanded to support restaurants, cafes, and food courts, making it a versatile solution for the hospitality industry.

IX. REQUIREMENTS

The MetaVisionn system requires both hardware and software components to function effectively. On the hardware side, a system with at least an Intel i3 processor, 4 GB RAM, and sufficient storage space is required, although higher specifications are recommended for better performance. A stable internet connection is essential for real-time voice processing and system communication. On the software side, the system can be developed and deployed on operating systems such as Windows, Linux, or macOS. Programming languages such as Python and JavaScript are used for development, along with backend frameworks like Flask, Django, or Node.js. The frontend is built using HTML, CSS, and JavaScript, while databases such as MySQL or Firebase are used for storing data. A modern web browser like Chrome, Edge, or Firefox is required to run the application interface.

X. CONCLUSION

MetaVisionn presents an innovative approach to modernizing food ordering systems in the hospitality industry through the integration of Artificial Intelligence and voice-based interaction. By enabling customers to place orders using natural speech, the system eliminates the need for manual input and significantly reduces communication errors. The use of real-time speech-to-text conversion and Natural Language Processing ensures accurate extraction of order details, while the display of both spoken and structured data enhances transparency and reliability. The system improves operational efficiency by streamlining communication between guests, hotel staff, and kitchen teams. Overall, MetaVisionn provides a fast, user-friendly, and intelligent solution that enhances customer experience and sets the foundation for future smart hospitality systems.

REFERENCES

[1] Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing*. Pearson.
[2] Zomato. (2024). Online Food Delivery Platform. Retrieved from https://www.zomato.com
[3] Swiggy. (2024). Food Ordering and Delivery System. Retrieved from https://www.swiggy.com
[4] Amazon Alexa. (2024). Voice Assistant Technology. Retrieved from https://developer.amazon.com/alexa
[5] Google Assistant. (2024). Voice Recognition System. Retrieved from https://assistant.google.com
[6] Rabiner, L., & Juang, B. (1993). *Fundamentals of Speech Recognition*. Prentice Hall.
[7] Chowdhury, G. G. (2003). Natural Language Processing. *Annual Review of Information Science and Technology*, 37(1), 51–89..