International Journal of Scientific Research Studies

ISSN (print): 3050-6905, ISSN (online): 3050-6913

DOI:10.58806/ijsrs.2026.v3i4n03

Volume 03 Issue 04 April 2026

Page No: 90-96

Design and Deployment of Speech-Enabled Real Time Automated Ordering System for Fastfood Restaurants in Africa.

Ufuoma Kazeem Okpeki[1], Edith Ugochi. Omede[2], Oyubu Akpovi Oyubu[3], Efenedo Gabriel Ilori[4], Emmanuel Ewere Obuseh[5]

[13,,4,5]Department of Electrical and Electronic Engineering, Delta State University Abraka, Oleh

[2]Department of Computer Science, Delta State University, Abraka, Nigeria.

ABSTRACT

Many online applications such as e-shopping applications with the aid of web technologies provide interactive and user friendly interfaces that enhance their usability. However, some intending users especially the visually impaired and those challenged in reading and writing are denied the full experience of the services provided by these applications. To assist in leveraging this difficulty, this paper has designed a framework for embedding voice-controlled system into a Restaurant Food ordering system which when fully iintegrated, will provide seamless interaction with the food ordering system as it allows both text and voice input. The standard structural system analysis methodology and the Unified Modelling Languages (UML) tools in Visual Paradigm have been used in the design. . The system was fully implemented and everything worked in line with designed specifications.

KEYWORDS: Speech, Enabled System, Online, Food Ordering, Rrestaurant

ARTICLE DETAILS

Published On:
06 April 2026

Available on:
https://ijsrs.org/

1.0 INTRODUCTION

Restaurant industry has greatly been impacted by Management Information System in various areas such as Online Menu Reservation, online ordering, inventory, and analytic. At the hit of Corona Virus pandemic, when there was restriction of movement and public gathering, restaurant operators lost a lot of customers and business fortunes. has stimulated the migration from conventional service procedure which involves waiters relating directly with customers to obtain their orders before processing to Automated Management system a necessity for many who desired to remain in the business (Georgescu etal, 2021). Though before now, food ordering system was entirely a manual process which involved waiters, pen and paper. In other words, restaurant services entail waiters to obtain customer ordering information in person and then take the orders to the kitchen for meal preparation (Edje, 2016). However, there are cases of some people especially the visually impaired and those with hands disabilities who are not able to experience the full serviceability of this application, thus embedding a voice controlled system into the application can enhance its usability by accommodating more users.

This research work proposes an enhancement to Real-Time Based Automated Ordering System which was a proffered solution system by (Sheryl, 2008)), to resolve the challenges of the conventional method used for ordering meals and beverages in fast food restaurants in Nigeria. This enhancement which this work tends to achieve by embedding Voice enabled system to the system will widen the user's scope. Many restaurants especially those in developed world make use of Table Reservation System (TRS) which enables the guests/customers to book for table space via internet website or the restaurant portals. This curbed the problem of long queue to book for table space either for dinner or any other occasion especially at peak time. The TRS helps restaurant management to serve their prospective clients better and manage cost more effectively. (Mandeep et al, 2018).

Though the comparative study of the existing SRS carried out by (Pahini, 2014) concluded that the performance of SRS systems are better than open source in terms of Phrase Recognition Rate (PRR) and Word Error Rate (WER), we decided to use Kaldi an open source SRS which has up to 84.9% word accuracy because it is readily available at no cost and is open for modifications and easily adaptable. Restaurant management system (RMS) has sort of customized set of words with which the open source SRS can easily be trained. (Georgescu et al, 2021).

Kaldi an open SRS was used to build the proposed system. The application was built using 3-tier architecture which consists of the client side (browser), the front-end (web app) and back-end (server and database). The client-side scripts (JavaScript and

HTML), was for presentation of information to the users. The back-end application was powered by server-side scripts (JavaScript and Python), and a database system, which allowed storing and retrieving of information and generating responses to send back to the client was built using a non-relational database tool, (MongoDB) that is capable of storing information in Binary JavaScript Object Notation (BJSON) which provides an efficient of an scalable application that supports different data type. The information with which the database will be fed was gotten from interactions with operators of several Fast food restaurants in Nigeria.

Speech recognition, or speech-to-text, is the ability of a machine to identify spoken words and convert it into readable text. Simple speech recognition software has a few vocabulary and may be able to identify only clearly spoken words and phrases. More sophisticated software can handle natural speech, different accents and various languages. (Raiyetunbi etal, 2020). Speech and Voice recognition are interchangeably used, though similar, they differ from each other. They are two different technologies and should not be confused. Voice Recognition (VR) is a biometric technology for identifying speaker's voice rather than the content of his speech while Speech Recognition (SR) is the process of converting spoken word into meaningful electrical signal. Meanings are assigned to the signals by comparing them with set of phonemic representations for a close match (Raiyetunbi et al, 2020). The phonemic representations are matched against predefined words in Word Vocabulary. Speech Recognition System (SRS) is birthed from Speech Recognition technology which aims at using natural language to trigger action. SRS enables more natural and effective communication among people and this often requires deep integration with various natural language processing (NLP) components. SRS can be used in many domains to break barriers in human-to-human and human-to-machine interactions. Speech recognition functions are built in various devices and text-focused programs to allow easy use of the device. Speech Recognition technology has made possible, a seamless human-machine interaction where people with visual impairments who may not be able to ordinary interact with computer/devices through keyboard and mouse can now do so at ease through use of voice. Many actually prefer the use of voice because of its easiness and convenience. It is more convenient to control your device through voice than using hand. This benefits of speech recognition technology is paving way for it in almost every domain. Tapping from this benefit, this research work aims at integrating SRS to Restaurant Management Information System focusing on Fast Food Restaurants in Nigeria. Another aspect of restaurant management touched is Inventory Control System(ICS). ICS minimizes wastage of food ingredients by automating the food stock inventory, keeping records of available stock, and automatically updating the present status of every ingredient in stock. This enables inventory control officer to know when to place order for any exhausted ingredient. (Narotama, 2013).Taxonomy of Speech recognition system was proposed by (Pahini, 2014), their work presents a voice controlled commodity e-commerce application using IBM Watson speech to text to demonstrate its usability. A real-time based automated Ordering System which is designed to enable prospective customers to place their meal order from their convenience without necessarily having to be present at the restaurant was proposed by (Narotama, 2013). The order is accepted using first come first serve (FCFS) device which will create orderliness on how the orders are received and attended to.

All these systems with their various contributions did not consider the enlargement of customers' coast by making provision for visually impaired individuals.

2.0. MATERIALS AND METHOD

The data used in this design workswere gotten from interactions with Fast food restaurant workers, academic journals and professional documents via internet. The research was carried out using, standard structured system analysis and design method (SSADM), which involved critical studying of previous works done on Restaurant management system and Speech Recognition System (SRS). The result of the study led to the identification of the lapses in existing Restaurant management system platforms which did not consider the visually impaired persons as prospective beneficiaries of the online facilities. The designs were carried out using Unified Modelling Language (UML) tools in Visual paradigm

2.1. Structure of Speech Recognition System (SRS)

The basic Structure of Speech Recognition System is shown in Figure 1.

Speech recognition systems must process, interpret and convert spoken words into machine/human understandable written language using computer algorithms. The input voice through a microphone is transmitted in form of wavelike electrical signals which is converted by the system's hardware into a digital signal. A speech recognition software analyses the signals to register phonemes by breaking down reliable context of human speech. The algorithms that process and organize audio into text are trained on various speech patterns, speaking styles, languages, dialects, accents and phrasings. The software also separates spoken words from background noise that often follows the signal.

Algorithms can also combine the predictions of acoustic and language models to offer outputs

Voice Input in form of raw speech which comes through microphone or telephone as high frequency signals which yields sequence of amplitude values over time.

Signal analysis. The Raw speech is simplified for subsequent processing by pre-processing it to extract useful features using any of the available signal analysis techniques. Among the most popular:

(i) **Fourier analysis (FFT)** yields discrete frequencies over time, which can be interpreted visually. Frequencies are often distributed using a *Mel* scale, which is linear in the low range but logarithmic in the high range, corresponding to physiological characteristics of the human ear.

(ii) Linear Predictive Coding (LPC) yields coefficients of a linear equation that approximate the recent history of the raw speech values.

The extracted features are moved into the search engine which is connected to Language and Acoustic models. Acoustic model consists of knowledge about acoustics, phonetics, variations in input source and environment, gender and dialect differences among different speakers while Language model consist of word structures, its semantics knowledge, possibly occurring words, similar words and likely sequence of occurrence. It also accommodates the semantics and functions related to a user's operation. Adaptation module is for modification of the output from acoustic and language modules.

2.2. Types of Speech Recognition:

It states, where the speaker is required to wait or pause between the utterances while the system accepts an utterance, processes it and gives corresponding result of the input speech.

(i) Connected Words: C the utterances of the words. Minimum pauses are identified as a unit which are trained as isolated word nodes. While the user waits, the system processes the input and generates corresponding output.

(ii) Continuous Speech: In continuous speech system, the users speak in a natural manner and the computer determines the content and is known as computer dictation. It is a difficult model to implement because the system needs special techniques to recognize text and at the same time determine the utterance boundaries.

(iii) Spontaneous Speech: Spontaneous speech basically can be thought of as natural speech which is not rehearsed. A spontaneous speech modelled system should be able to handle multiple natural speech. Speech recognition has become a complex and a challenging task because of the variability of the signals. Speech recognition requires certain definitions for its proper understanding. Speech recognition systems can be separated in several different classes by describing the types of utterances that they can recognize. Utterance is the speaking of a word or words that defines a single meaning to the computer. Utterances can be a single word, multiple words, a sentence, or even multiple sentences.

2.3 The Designed System

The system is modelled to allow voice or text input. The voice input is sent through microphone, which will be accepted, analyzed and synthesized by SRS to generate corresponding texts, which triggers the server system to respond accordingly and sends messages to Service department (kitchen and Sales personnel), who in turn acts in response to received request. The feedback will also be converted from text to speech back to the customer/client. The hardware structure of the designed system is shown in Figure 1.

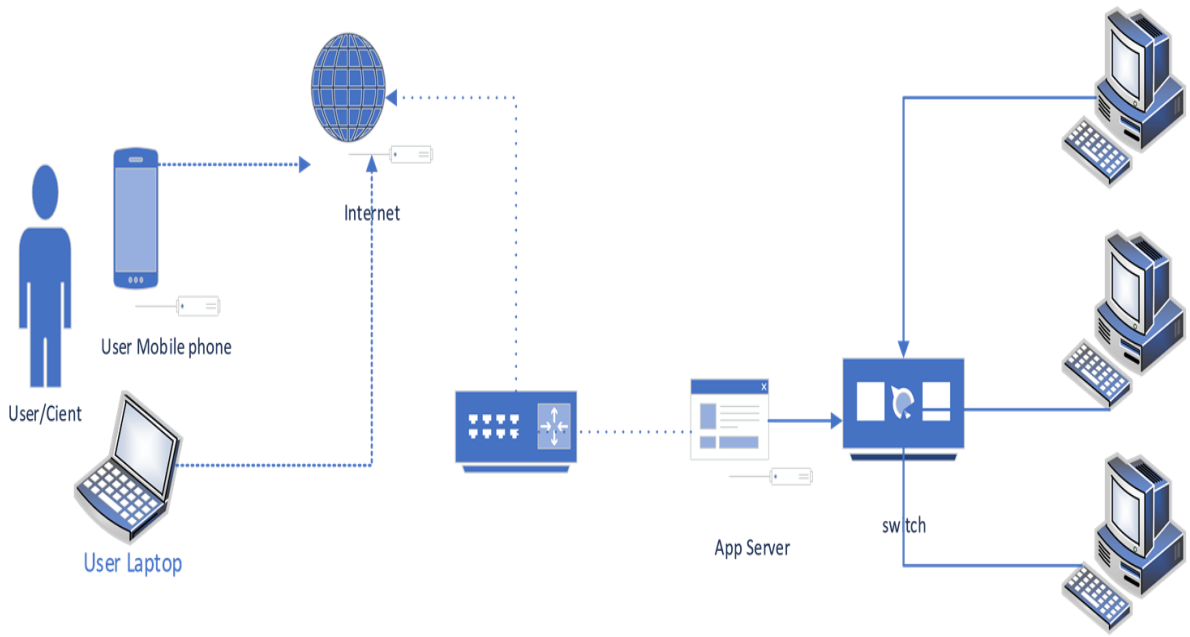


Figure 1: The Hardware structure of System

2.4 The Use case diagram

The use case diagram provides the graphical overview of the goals a user desire to achieve using the designed system. The Figure 2 below depicts that the customer has to launch the browser and connect to the restaurant through its web address. On the restaurants web page, the customer views/listen to available services and makes order. He can also make payments and get responses back from the system. While the admin on the other hand should update the system with available restaurant services.

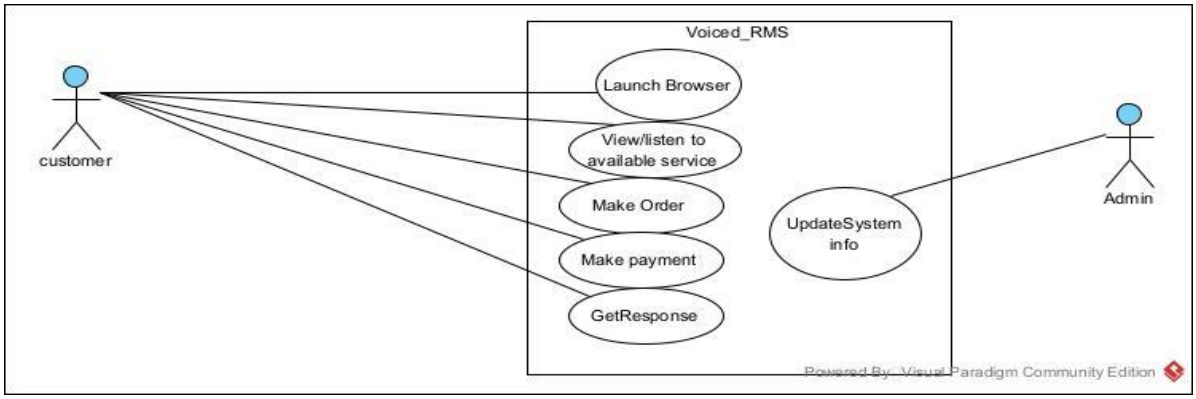


Figure 2: System Use case diagram

2.5. The system Activity Diagram

The activity diagram is used to represent the series of actions and flow of control in the designed system. It involves the input modulethrough which input which, can be text or speech is sent to the system. For text input, the processing of the request is carried out by RMS App, and corresponding response is sent to the customer, but if the input is speech, Speech to corresponding Text (STT) service is invoked for speech analysis that converts the speech into corresponding text which if matches with keywords in the database, the request is processed. The Text to Speech (TTS) service of SRS is invoked for re-conversion of the text to speech and voice response corresponding to request made, is sent to the customer (Itunuoluwa et al, 2014). The activities are as shown in Figure 3.

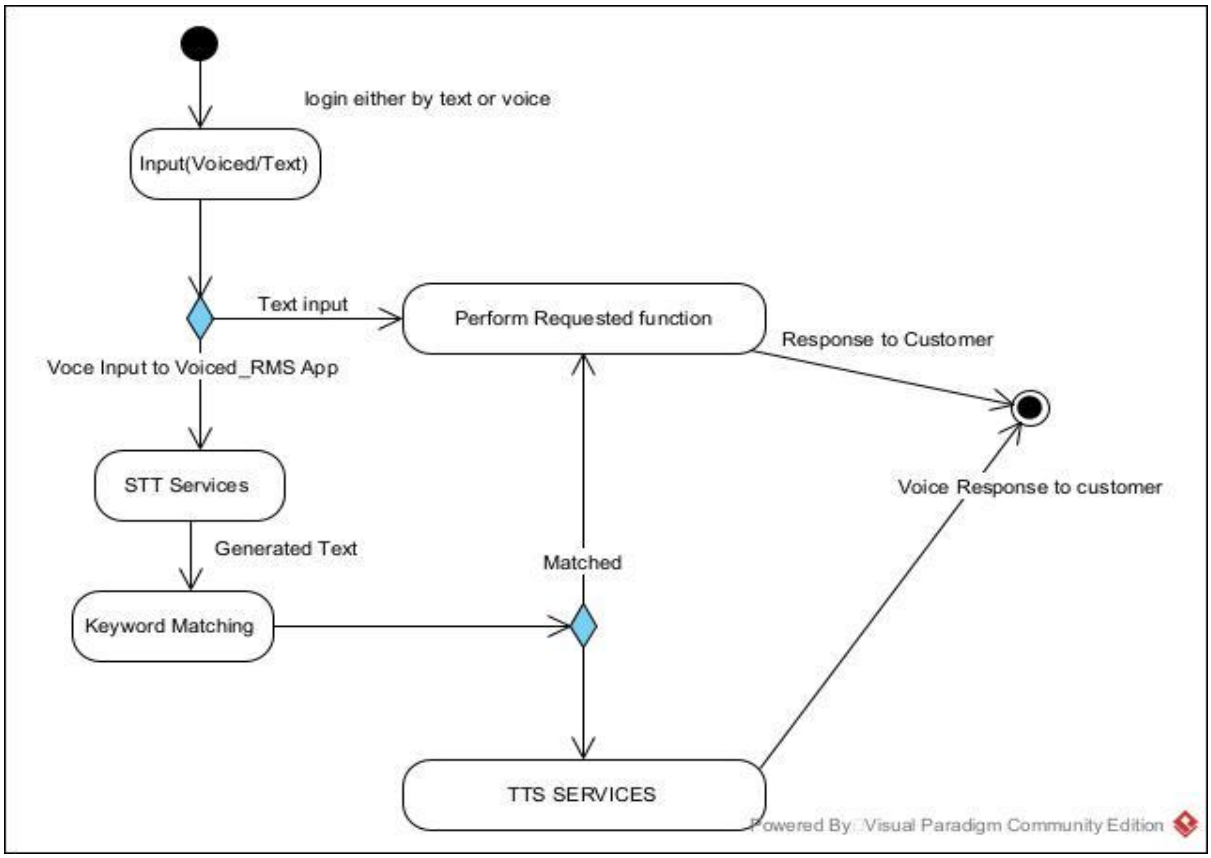


Figure 3: Activity diagram of the designed System

2.6. The Designed System Sequence diagram.

Sequence diagram in Figure 4 describes how the different component of the designed system interacts to fulfil the required use case goal. The User launches the web app by typing or calling the web address of the restaurant, and makes a request either using text/voice. The voice input is analyzed by STT service of incorporated SRS which converts it to text, the converted text is matched with keywords in the database, if match is found, the server forwards the request to the switch which in turn dispatches the request to appropriate service department (Sales/Kitchen). The request is responded to and is forward back to the server through the switch. TTS service of SRS is invoked to convert the response from text to speech before it will be forwarded to the customer.

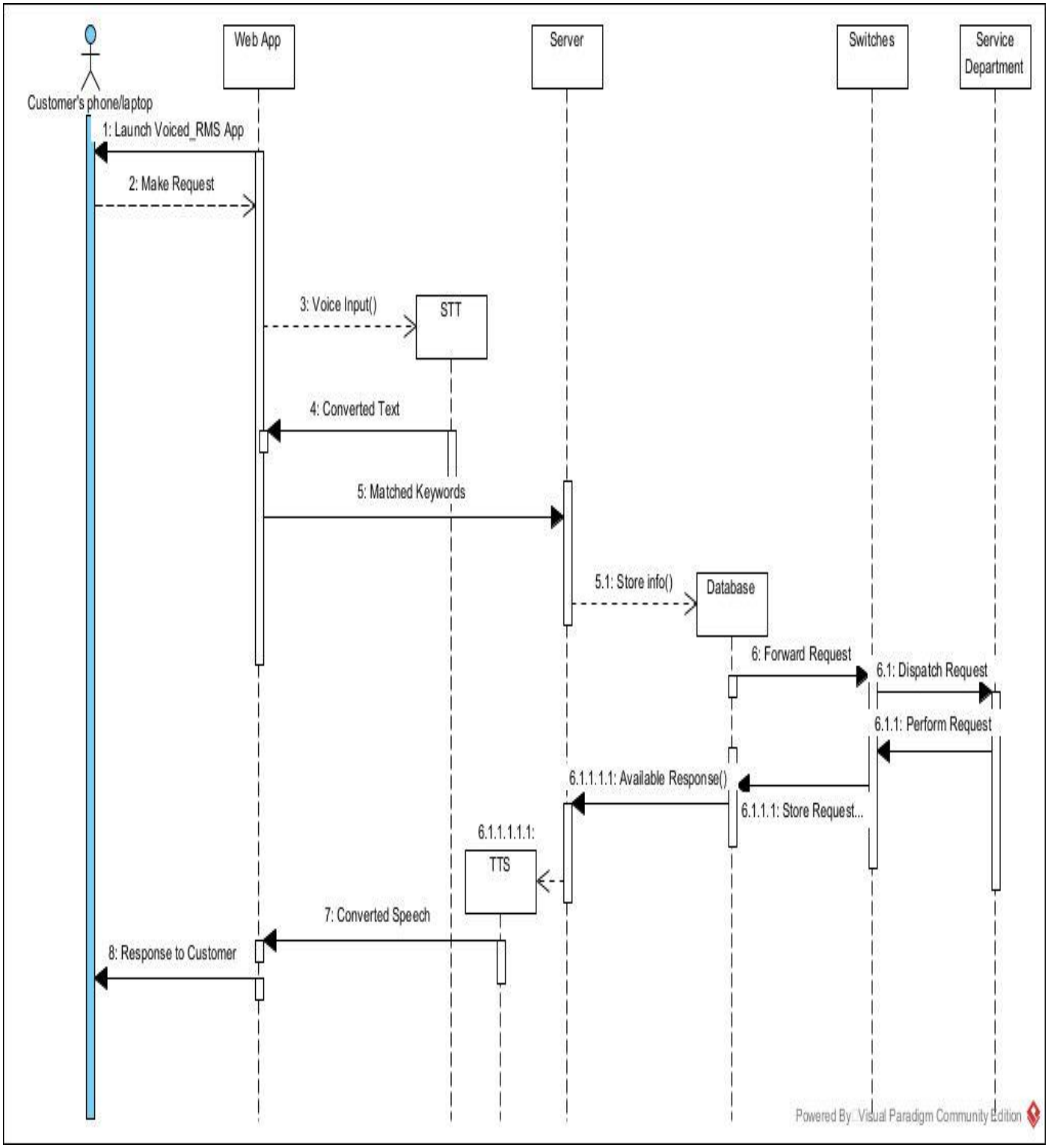


Figure 4: Sequence Diagram

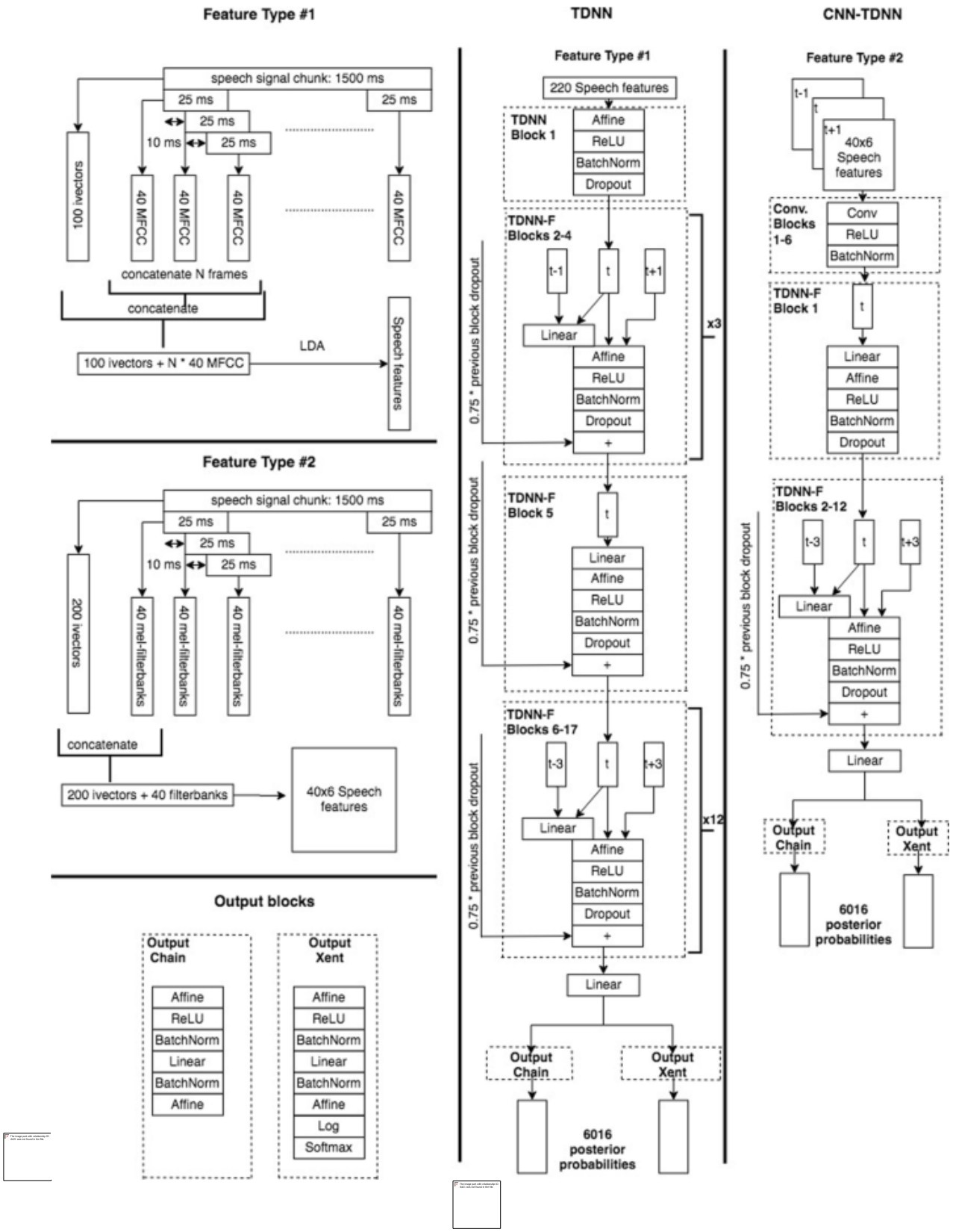


Figure 5. Kaidi full circuit diagram

Figure 6 is the full circuit diagram for kaldi TTS and STT implementation platform.Isa multi-facet system, made of based acoustic model, a phonetic model and a language model, all these made up the core components of the pipeline system.

3.0. RESULTS AND DISCUSSION

The designed system was tested and everything worked in line with designed specifications. if the system is widely implemented or deployed, thedesigned will not only reduce the stress of long queues in restaurant but also widens the scope of the potential customers by providing opportunity for visually impaired persons to also be beneficiaries of online facilities.

4.0 CONCLUSION

The designed system which is an enhancement on existing system will not only improve customers' ordering experience but will also widens the scope of potential customers by extending its usability to visually impaired and those that have challenges with reading and writing thereby increasing the restaurant's sales income. The SRS enable web application will promote an ease of interaction and creates a platform where users can easily transact business using their natural voice without stress of hand and eyes on their devices.

For future work, the application can be improved upon by incorporating a language translator API, which will enable usersto place orders in their own native languages thereby removing the embargo placed on those who cannot use web because of inability to understand convectional web language.

Acknowledgement

We hereby acknowledged late Engr. Prof. Clement Ojeabu of the department of electrical/electronic engineering, Edo State University Ekpoma, for proof reading the work before his demise. May his soul rest in the bosom of the Almighty Allah.

Declaration of Interest Statement

We the authors hereby states that we have no financial commitment with no body.

REFERENCES

1) Edje E.A (2016):The design of a real time based automated ordering system for fast-food restaurants
2) Georgescu, AL., Pappalardo, A., Cucu, H.(2021). "Performance vs. hardware requirements in state-of-the-art automatic speech recognition". J AUDIO SPEECH MUSIC PROC. 2021, 28 . https://doi.org/10.1186/s13636-021-00217-4
3) Narotama (2013). "Material: Inventory Control", http://dosen.narotama.ac.id/wp-content/uploads/2013/02/Chapter-14-Materials-Inventory-Control.pdf, in Edje(2016)
4) Pahini A. Trivedi(2014). Introduction to Various Algorithms of Speech Recognition: Hidden Markov Model, Dynamic Time Warping and Artificial Neural Networks. © 2014 Raiyetunbi, Oladimeji Jude and Ayeh Emmanuel (2020). An Interactive Cloud Based User Oriented, Dynamic and Intelligent Text-To-Speech Module. *East African Scholars J Eng Comput Sci*; Vol-3, Iss- 1 (Jan, 2020): pg. 1-11
5) Sheryl L. Kimes (2008). "The Role of Technology in Restaurant Revenue Management", http://scholarship.sha.cornell.edu/cgi/viewcontent.cgi?article=1412&context=articles,