VOICE BASED BILLING SYSTEM

D. RANJITH
B.E. Department of CSE, Sri Ramakrishna Engineering College, Coimbatore,
Tamil Nadu, India

G. SRUTHI
B.E. Department of CSE, Sri Ramakrishna Engineering College, Coimbatore,
Tamil Nadu, India

S. SWATHILAKSHMI
B.E. Department of CSE, Sri Ramakrishna Engineering College, Coimbatore,
Tamil Nadu, India

Mrs. N. VINOTHA
Assistant Professor (Sr.G), Department of CSE, Sri Ramakrishna Engineering College, Coimbatore,
Tamil Nadu, India

---***---

Abstract - *Recent advances in speech recognition and natural language processing have led to a new generation of voice-based interfaces. Speech is the first important primary need and the most convenient means of communication between people. The communication among human computer interaction is called human computer interface. The speech input can be used in varying domains such as automatic reader and for inputting data to the system. Speech recognition can minimize the use of text and other types of input, at the same time minimizing the calculation needed for the process. Decade back speech recognition was difficult to use in any system, but with elevation in technology leading to new algorithms, techniques and advanced tools. Now it is possible to generate the desired speech recognition output. This proposed system gets the input through voice which are the items purchased by the customer, that will be converted into text. To the converted text, certain NLP methods are applied to tokenize the words or sentences. Once the words are tokenized it proceeds with the calculation of bill.*

*Key Words***: SPEECH RECOGNITION, NLP, HUMAN COMPUTER INTERFACE**

1. INTRODUCTION

The Speech recognition is the interdisciplinary sub-field of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers. It is also known as automatic speech recognition (ASR), computer speech recognition, or just speech to text (STT). The STT module aims to provide a user- friendly application to general users. The application provides a multi-functionality platform for users to communicate, listen or narrate conveniently. The speech to text mode converts an inputted voice to text through Google API servers which gets displayed and given as an input to next module.

Most of the existing systems just converts the speech into text and stores it in a database where user can upload or retrieve the data's in future if needed. This application not only just converts the speech to text but also performs tokenization of words which is one of the most common tasks in Natural Language Processing (NLP).

Natural Language Processing is a way for computers to analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation. Also along with the speech transcription system, it also calculates the bill for items purchased by the consumer and the generated bill gets stored in a .txt file in the vendor's system. To use the speech recognition, the user has to be loud and clear so that the command is properly executed by the system.

2. RELATED WORK

A Voice to Text transcription system [1] for NGO's was built to record the conversation during surveys and convert it into text and save it. This system includes an open source application. The CMU Sphinx toolkit was used for speech recognition. The system supports multi-language recognition. The CMU Sphinx toolkit utilizes acoustic model, phonetic dictionary and language model. The user had recorded his/her voice through the mobile application then recognition and transcription was done through CMU Sphinx toolkit. The transcription file was saved as a text file in the device memory which user can upload and retrieve data from the database server through the application. The system could recognize and translate multi language speech or voice. The application reduces the time wasted by the counselor to manually type the whole session to text.

An interactive database system for automatic speech recognition [2], in which speech was recognized using speech

recognizer and the recognized voice was converted into text. An xml configuration file for the created database was written. The configuration file was taken as an input for the conversion of text into SQL query. Then text was converted into a standard SQL query. For the constructed SQL query necessary details were retrieved from the database. This process was repeated until the user was satisfied with the retrieval results.

An Intelligent Hands Free Speech based SMS System on Android[3] in which Voice or signaled input was inserted through any speech device such as microphone, then speech was processed and converted to text hence able to send SMS, also Phone number was entered either by voice or by selecting it from contact list. Voice had opened up data input for a variety of user's such as illiterate, handicapped, as if the person cannot write then the speech input was a boon and other's too which lead to better usage of the application. Users who have no knowledge about the SQL queries can also use this for retrieving the details from the database.

A Text to Speech conversion module [4] in which Text converted to Speech, Text file converted to Speech, Text in various Languages converted to Speech, Image converted to Text and Image converted to Speech using MATLAB as a programming tool. The various methods used were Preprocessing, Unicode Conversion, Segmentation, Concatenation, Prosody and Smoothing, to be then combined in an application for easy access and usability. The text-to-speech mode converted a text file or inputted text to speech which then was narrated/read using the voice database used by Microsoft SAPI.

A querying database system using a voice-based interface called EchoQuery [5] was developed. It showed the practicality and utility of QbV for relational DBMS using a proof-of-concept system called EchoQuery. NLIDB had been very limited since they mainly focused on constructing interfaces for individual domains and not general purpose interfaces for exploring arbitrary data sets. They used deep-learning models to allow for a robust translation from natural language to SQL. The query interface of EchoQuery is inspired by regular human-to-human conversations where the system can ask questions back to the user.

3. OVERVIEW OF PROPOSED SYSTEM

A Voice based billing system has been proposed to do the billing for the items purchased which gets voice as an input. The input voice is either a word or a sentence which gives the name of the items purchased and its quantity. Word tokenization – the process of splitting a large sample of text into words. This is a requirement in natural language processing tasks where each word needs to be captured and subjected to further analysis like classifying and counting them for a particular sentiment etc. The output of word tokenization can be converted to data frame for better text understanding in machine learning applications. Then the

converted texts are given as input for the calculation of bill. Once the bill has been calculated, the text file will be created with the calculated output that can be viewed. The execution time is also calculated and displayed. In addition to it, the accuracy for the converted text is calculated using confusion matrix and the graph has been plotted.

LIST OF MODULES

A. Speech to Text
B. Word Tokenization
C. Bill Generation

A. SPEECH TO TEXT

Speech-to-Text can use one of several machine learning models to transcribe your audio file. Google has trained these speech recognition models for specific audio types and sources. When you send an audio transcription request to Speech-to-Text, you can improve the results that you receive by specifying the source of the original audio. This allows the Speech-to-Text API to process your audio files using a machine learning model trained to recognize speech audio from that particular type of source. To specify a model for speech recognition, include the model field in the object for your request, specifying the model that have been used.
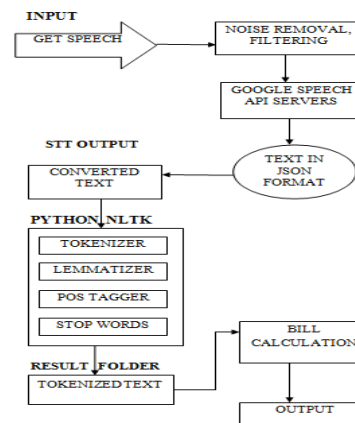

Fig 1 : Block diagram of proposed system

B. WORD TOKENIZATION

One common task in NLP (Natural Language Processing) is tokenization. "Tokens" are usually individual words and "tokenization" is taking a text or set of text and breaking it up into its individual words. It is the process of breaking up the given text into units called tokens. These tokens may be words or number or punctuation mark. Tokenization does this task by locating word boundaries. Ending point of a word and beginning of the next word is called word boundaries. Tokenization is also known as word segmentation. The method word tokenize() is used to split a sentence into words. The output of word tokenization can be converted to data frame for better text understanding in machine learning applications. It can also be provided as

input for further text cleaning steps such as punctuation removal, numeric character removal or stemming. Word tokenization becomes a crucial part of the text to numeric data conversion.

C. BILL GENERATION

Once the words and sentences are tokenized, the tokenized words are given as the input for the bill calculation. The input contains the item purchased and its quantity. The consumer can give the items purchased through voice, and consumer can also cancel the purchased items before it gets calculated. The bill gets calculated once the system gets input as the keyword TOTAL. Once the keyword total is given as an input the bill gets generated and gets displayed in the python shell. It is also displayed in the web page with the bill number, date and time.

4. OPEN SOURCE TOOLS AND PYTHON LIBRARIES

PyAudio

PyAudio provides Python bindings for PortAudio v19, the cross-platform audio I/O library. With PyAudio, you can easily use Python to play and record audio streams on a variety of platforms (e.g., GNU/Linux Microsoft Windows, and Mac OS X).

NLTK

The Natural Language Toolkit, or more commonly NLTK, is a suite of libraries and programs for symbolic and statistical natural language processing (NLP) for English written in the Python programming language. It was developed by Steven Bird and Edward Loper. It includes graphical demonstrations and sample data. It is intended to support research and teaching in NLP or closely related areas, including empirical linguistics, cognitive science, artificial intelligence, information retrieval, and machine learning. There are 32 universities in the US and 25 countries using NLTK in their courses. It supports classification, tokenization, stemming, tagging, parsing and semantic reasoning functionalities.

XAMPP

XAMPP is a free and open-source cross-platform web server solution stack package developed by Apache Friends, consisting mainly of the Apache HTTP Server, MariaDB database, and interpreters for scripts written in the PHP and Perl programming languages.

5. EXPERIMENTAL RESULTS

The system has been tested for certain fruits and undergoes a STT method and NLP methods which converts the speech input into text and calculates the billing amount with the execution time being displayed. The main objective of this system is to reduce the time complexity and to improve the accuracy than the existing system. The accuracy is calculated for the converted text using confusion matrix and compared with the existing system's accuracy.

Classification Rate or Accuracy is given by the relation:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

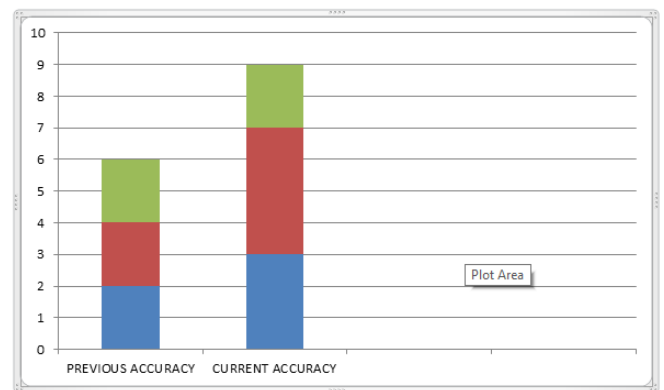The accuracy of the proposed system is 90% respectively.

Chart – 1 Comparison of accuracy

6. CONCLUSION

The developed system will transcribe speech i.e spoken words or conversation into text and store all relevant data. Vendor's use this data for the calculation of bill for the items purchased by the consumer. One of the main objectives is to convert the inputted speech into text even it has a noisy background. The system is implemented for both the types of speech input i.e it can be either a name of the fruit along with its quantity or a sentence which contains the fruit name and the purchased quantity.

Implementation of deep learning techniques for the conversation part can provide completely different results. Another option is to update the available stock and its respective price through voice. As of now a voice based billing system has been built and in future certain Machine learning techniques can be applied to the same to improve its efficiency.

7. REFERENCES

[1] Ashfaq Shaikh, Burhanuddin Lakdawala, Farhan Khan, "Voice to Text transcription using CM Sphinx", Second International Conference on Inventive Communication and Computational Technologies (ICICCT 2018), IEEE Xplore.

[2]S.Nareshkumar, N.Mariappan, K.Thirumoorthy, "Database Interaction Using Automatic Speech Recognition", International Journal of Innovative Research in Science Engineering and Technology, Vol.3, March 2014.

[3] GulbaksheeDharmal, Dr. Vilas Thakare, Dr. Dipti D. Pati, "Intelligent Hands Free Speech based SMS System on Android", International Conference on Advances in Human Machine Interaction, March 03-05, 2016, IEEE Xplore.

[4] Hussain Rangoonwala, Vishal Kaushik et al , "Text To Speech Conversion Module", IEEE Transaction, Volume 115 No. 6 2017.

[5] PrasetyaUtama, Nathaniel Weir et al, "Voice-based Data Exploration: Chatting with your Database", Springer-2011, 114 p.

[6] Qiang Fu, Yong Zhao, Member, IEEE, and Biing-Hwang Juang, "Automatic speech recognition based on non- uniform error criteria", IEEE Trans. On audio, speech, and language processing, vol. 20, no. 3, march 2012.

[7] Q. Fu, D. Mansjur, and B.-H.Juang, "Empirical system learning for statistical pattern recognition with non-uniform error criteria", IEEE Trans. Signal Process., vol. 58, no. 9, pp. 4621–4633, Sep. 2010.

[8] F. Li and H. V. Jagadish, "Constructing an interactive natural language interface for relational databases", PVLDB, 8(1):73–84, 2014.

[9] G. Lyons, V. Tran, C. Binnig, U. Çetintemel, and T.Kraska, " Making the case for query-by-voice with echoquery", International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016, pages 2129–2132, 2016.