IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Smart Ordering System using Speech Recognition

[1]Sandhya R, [2]Sabarinathan S, [3]Swathi V, [4]Dr. M. Pandi
[1, 2, 3] BE Student, [4]Assistant Professor (SG)
[1,2,3,4] Department of Computer Science and Engineering
[1,2,3,4] Dr. Mahalingam College of Engineering and Technology, Pollachi, India

Abstract: **In the era of Information technology human tend to develop better and more convenient lifestyle. There is no time for people to visit grocery shop to buy products. Simply they make a phone call and order the products they want. But, sometimes it may be a set back as if the call gets engage. Rather it will be a better option to appoint a person to manage the phone call. Our project focuses on developing a system which helps people making an order through an android application using speech recognition to reduce the manpower in grocery shop and to reduce waiting time of the customer. This system records the speech and converts the speech into text. The converted text is processed by Naïve Bayes classification to obtain suitable intents such as product name and quantities. Finally the intent is processed and the corresponding orders are placed and the invoice will be generated to the customer.**

Index Terms **- CNN, intent, speech recognition, text classification.**

I. INTRODUCTION

Every day a Smartphone user may look for a new application dedicated for his/her need. Android makes it easier for consumers to get and use new content and applications on their Smartphone. For the past several decades, designers have processed speech for a wide variety of applications ranging from mobile communications to automatic reading machines. Speech recognition reduces the overhead caused by alternate communication methods. Speech has not been used much in the field of electronics and computers due to the complexity and variety of speech signals and sounds. However, with modern processes, algorithms, and methods we can process speech signals easily and recognize the text. In this project, we are going to develop an android application for speech recognition. This application presents a really on-demand, fast and user friendly android application speech recognition, i.e. by using this application it will achieve more work with less time and do many at that time. For that one of the best applications is the smart ordering system using speech to text conversion.

Developing a speech-enabled applications for mobile devices typically requires comprehensive analysis of several speech processing engines and different architectural approaches before the desired application can be built. In addition, framework and tool support is often insufficient, which makes developing applications difficult. Developing a speech based application for mobile devices requires more work, since speech recognition systems and mobile devices vary dramatically in their capabilities. While mobile devices can concisely be classified by their processing power, memory, operating system and wireless network speed it is a bit trickier for speech recognition engines.

This project presents a comprehensive approach that comprises a profound classification of speech recognition systems for mobile applications and a framework for mobile and distributed speech recognition. We are developing such an application which is solely useful for user by taking speech as input and converting it into text. The customers can interact with the application through speech commands which in turn get interpreted and then decodes the analog signal using spectrum analyzer then the command will be transformed into a text string which will be utilized for further process. This is very useful for the user for making many work at a time and your time will be save with more output.

II. LITERATURE REVIEW

Shivakumar K M, Varsha V Jain and Krishna Priya P [1] proposed the speech to text conversion of Kannada general conversational sentences using CMU Sphinx speech processing system.An acoustic model is created using Sphinx training tool.In language model probability of unigram bigram and trigrams of corpus is calculated and used by decoder to predict the possible combination of words and phrase. The experiment shows high accuracy for randomly selected sentences compared to sequential sentences.

N.Uma Maheswari, A.P.Kabilan, R.Venkatesh [2] proposes speaker independent speech recognition system based on phoneme identification" A bottom-up approach, in which a speech signal is resolved into a string of phonemes and phoneme recognition forms the basis for word recognition which in turn constitutes the full text output. Phonemes are recognized using artificial neural network and subsequently words are recognized from the clusters of phonemes using HMM. A Neural Network with trained delays and widths and random weights classifies 96% of the phonemes correctly.

P. Saini, P. Kaur, and M. Dua, [3] Speech recognition system for Hindi language using Hidden Markov Model is used to develop the system. It identifies the isolated words using acoustic word model. When the amount of state lessens and the word length goes less, it is then the system performance gets better. When 10 states in HMM topology are taken the current system gives the accuracy of 96.61% and 95.49% with word error rate 6% and 8%.

Has¸im Sak, Andrew Senior, Kanishka Rao, Franc¸oise Beaufays, [4] proposes Fast and Accurate Recurrent Neural Network Acoustic Models for Speech Recognition. Long Short-Term Memory (LSTM) RNN acoustic models for large vocabulary speech recognition. The systems use cross-entropy training with HMM context dependent (CD) state targets followed by sequence training. Connectionist Temporal Classification (CTC) models use a "blank" symbol between phonetic labels and propose an alternative loss to conventional cross-entropy training. Trained CTC models for acoustic speech recognition show that with appropriate features and the introduction of context dependent phone models they outperform the conventional LSTM RNN models by 8% relative in recognition accuracy.

Titouan Parcollet, Ying Zhang, Mohamed Morchid, Chiheb Trabelsi, Georges Linarès,Renato De Mori and Yoshua Bengio,[5] proposes Quaternion Convolutional Neural Networks for End-to-End Automatic Speech Recognition. Sequence-to-sequence mapping with the CTC model using Quaternion CNN. QCNNs obtain a lower phoneme error rate (PER) with less learning parameters than a competing model based on real-valued CNNs. This system product the previous output with current input in order to get low error rate. It results PER of 19:64% is obtained with just a QCNN without HMMs, RNNs, attention mechanism, batch normalization, phoneme language model, acoustic data normalization or adaptation.

III. PROPOSED SYSTEM

The system records the speech with the help of mic in the system and converts the speech into text through speech recognizer. The converted text is processed by Natural Language Processing Technique (Naïve Bayes classification) to obtain suitable intents such as product name and quantities. The classifier uses the Bayes Theorem. It predicts probabilities for each intent in given record. The intent with the highest probability is considered as the most likely intent. Finally the intent is processed and the corresponding orders are placed and the invoice will be generated to the customer.

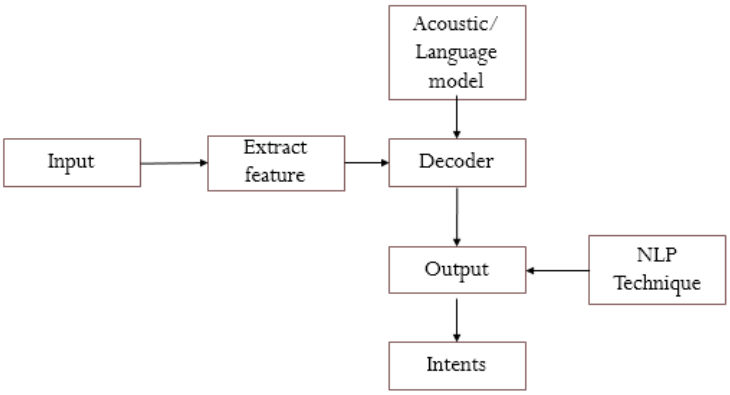

Figure 1 Block Diagram of proposed system.

- Input signal- Voice input by the user.
- Feature Extraction- It should retain useful information of the signal, deduct redundant and unwanted information, show less variation from one speaking environment to another, and occur normally and naturally in speech.
- Acoustic model- It contains statistical representations of each distinct sounds that makes up a word.
- Decoder- It will decode the input signal after feature extraction and will show the desired output.

- Language model- It assigns a probability to a sequence of words by means of a probability distribution.
- Output- Interpreted text is given by the computer.

IV. METHODOLOGY

The smart ordering system gets the input from user as audio through microphones. It will employ natural language processing techniques to understand the user's request and then convert it into text. The system consists of two main modules admin and user. Admin can login and use dashboard, view products and customer details and updating invoice. User can login or sign up for new registration, search product through voice based commands and get notification of invoice.

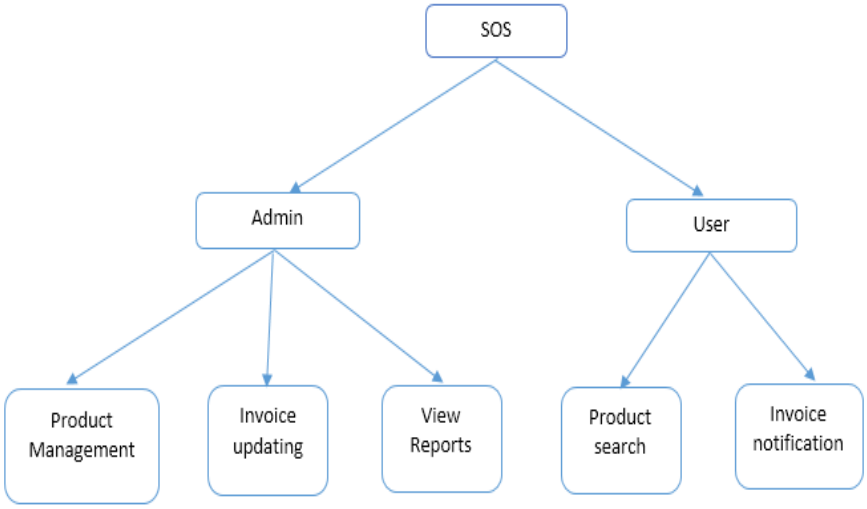

Figure 2 Flow diagram of Smart Ordering System

4.1 Registration

First, user should able to register their details for experiencing this application. Once the registration is successful user can get and user id which can be useful for user login, tracking of customer order by admin. User can login using their user id and password.

4.2 User Operations

Product Search: After successful login, user can able to search the product available in database using voice based commands. First user has to speak the product and their quantity in the respective fields. User can also able to edit the list. The products get added into the cart while speaking. The speech is converted into text using Speech Recognition package in android. Once the order is placed user can get an invoice of the product.

Invoice: Once the order is placed, user can get an invoice for the product. Invoice contains transaction id, name of the product, quantity and amount of the product if the products are available. In case the product is not available user can a notification of unavailable products.

4.3 Admin Operations

Product Management: Admin can add product and their quantity to the database after successful login using their unique names and password.

Invoice Processing: After receiving the orders from the customer, the system will analysis the order number, product description, quantity required behalf of admin. Based on the customer's requirement, the system will calculate the product amount and revert back for customer's view. In case the product is not available the notification is send to both the customer and admin for future use.

View Reports: Admin can view the customer details who all are registered and make orders, admin can also able to view the feature products that is unavailable products in the database.

V. RESULTS

The performance metrics for speech recognition commonly is word error rate (WER). The general difficulty of measure speech recognition performance lies in the recognized word sequences could have different length from the reference word. Speech recognition word error rate (WER) is a good tool for comparing different system and for evaluating improvements with a system. The WER still effective for speech recognition where errors can be corrected by typing, such as, dictation. However, for almost any other type of speech recognition systems, where the goal is more than transcription, it is necessary to look for an alternative, or additional, evaluation framework.

When evaluating performance of a speech recognition system, the word accuracy (WAcc) is used instead of word error rate (WER). The Word Accuracy (WAcc) is a metric used to evaluate speech recognizers.

$$WAcc = 1 - WER = \frac{N - S - D - I}{N} \quad (5.1)$$

where,

S for the number of substitutions,
D for the number of deletions,
I for the number of insertions,
N for the number of words in the reference.

The Word Error Rate (WER) and average Word Accuracy (WAcc) for various words such as rice, sugar, bread, jam and ghee is calculated for n rounds and by different persons. The results are tabulated.

Table1 Accuracy of words spoken by different person

Person	Rice	Sugar	Ghee	Bread	Jam
1	90	95	88	93	89
2	85	94	90	95	90
3	83	97	87	97	92
4	95	98	95	92	87
5	89	92	94	91	95
Average (%)	87	96	90	94	91

The average accuracy of each word spoken by different person. It is seen that the word rice has less accuracy than other words. Because, the word rice sounds like raise, nice etc. But other than this all other words are converted accurately in short period of time.

VI. CONCLUSION

The speech recognition is bringing the convenience to the public, to make public have the chance to experience this awesome technology and gain the advantages from the "Smart Ordering System using Speech Recognition". Thus the improvisation of the product ordering, order processing, bill generation, feature product collection is done with the aid of Android platform. In our experiment, the accuracy can be increased up to 96%. With the offline speech recognition system, the project application's performance is enhanced to a certain level, which will be lesser user undesired recognized result being processed. The future work is planned to implement the model of speech recognition for different language.

REFERENCES

[1] Titouan Parcollet, Ying Zhang, Mohamed Morchid, Chiheb Trabelsi, Georges Linarès,Renato De Mori and Yoshua Bengio," Quaternion Convolutional Neural Networks for End-to-End Automatic Speech Recognition", arXiv:1806.07789v1 [cs.SD] 20 Jun 2018.

[2] Dimitri Palaz, Mathew Magimai.-Doss, Ronan Collobert," Analysis of CNN-based Speech Recognition System using Raw Speech as Input", *INTERSPEECH 2015* .

[3] B N.Uma Maheswari, A.P.Kabilan, R.Venkatesh," Speaker independent speech recognition system based on phoneme identification",*IEEE International Conference on Computing, Communication and Networking (ICCCN 2008)* 978-1-4244-3595,2018.

[4] Has¸im Sak, Andrew Senior, Kanishka Rao, Franc¸oise Beaufays," Fast and Accurate Recurrent Neural Network Acoustic Models for Speech Recognition", tararXiv: 1507.06947v1 [cs.CL] 24 Jul 2015.

[5] Debatin, L., Filho, A. and Dazzi, R.,"Offline Speech Recognition Development - A Systematic Review of the Literature."in *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018) - Volume 2*, pages 551-558 ISBN: 978-989-758-298-1.

[6] "WordErrorRate"Wikipedia,[Online].Available: http://en.wikipedia.org/wiki/Word_error_rate [Accessed September 2019].

[7] X. Zhang, Q. Xiong, Y. Dai and X. Xu, "Voice Biometric Identity Authentication System Based on Android Smart Phone," *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*, Chengdu, China, 2018, pp. 1440-1444.

[8] Reeta Bandhu et al. *"*Offline speech recognition on android device based on supervised learning,"*International Journal of Advance Research, Ideas and Innovations in Technology*.2019.

[9] Wang, N. (2018). "Application of Android System Software in English Listening and Speaking Teaching." *Educational Sciences: Theory & Practice, 18(5),* 1846-1855.

[10] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Gerald Penn," Applying convolutional neural networks concepts to hybrid nn-hmm Model for speech recognition," Google Scholar 2015.

[11] Shivakumar K M , Varsha V Jain and Krishna Priya P, *"*A study on impact of Language Model in improving the accuracy of Speech to Text Conversion System*," IEEE International Conference on Communication and Signal Processing,* April 6-8, 2017, India 978-1-5090-3800-8/17/$31.00.

[12] P. Saini, P. Kaur, and M. Dua, "Hindi Automatic Speech Recognition Using HTK," Int. J. Eng. Trends Technol., vol. 4, no.6, pp. 2223–2229,June 2017.

[13] "TIMIT" Wikipedia, [Online]. Available: https://en.wikipedia.org/wiki/TIMIT [Accessed September 2019].