International Conference on Advances in Human Machine Interaction (HMI - 2016), March 03-05, 2016, R. L. Jalappa Institute of Technology, Doddaballapur, Bangalore, India

Intelligent Hands Free Speech based SMS System on Android

Gulbakshee Dharmale[1], Dr. Vilas Thakare[3], Dr. Dipti D. Patil[2]
[1,3] Computer Science Dept., SGB Amravati University, Amravati, INDIA.
[2] Computer Engineering Dept., MKSSS's CCOEW, Savitribai Phule, Pune University, Pune, INDIA.

Abstract--**Over the years speech recognition has taken the market. 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. A 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. One such method is the hidden markov models which is used in this paper. Voice or signaled input is inserted through any speech device such as microphone, then speech can be processed and convert it to text hence able to send SMS, also Phone number can be entering either by voice or you may select it from contact list. Voice has opened up data input for a variety of user's such as illiterate, handicapped, as if the person cannot write then the speech input is a boon and other's too which can lead to better usage of the application.**

Keywords: HMM; Speech Recognition (SR); Android;

I. INTRODUCTION

The use of voice in mobile phones has opened up the market for voice application for variety of the user's such as handicap etc. Similar kind of technology was apple's well known siri. The siri enable the user's to call contacts, send email and many more functionality. Similar to it was google speech recognition (SR) api. The major difference between both of them is that siri can understand multiple phrases and words but google voice majorly focused for specific phrases and keywords [1]. Speech recognition adds another dimension to the classic keyboard input which leads to ease of the user side i.e. manipulation of text is far easier than the classic method. This application uses the google api which uses the hidden markov models (HMM) method to send sms, In this sms application the user has to speak the numeric characters as the contact information on which sms has to send and message for the receiver.

This application also included that user can only input numeric character for contact information, i.e. the security validation for number is done. SR will listen to input and convert numeric to text and will be displayed on contact information to verify. If any user try to insert any other character into the information an error would be displayed e.g. if user speaks his name for contact, it will be displayed as invalid contact. The message box can accept any character. To use the speech recognition user has to be loud and clear so that command is properly executed by the system.

II. SPEECH RECOGNITION

The system shown here will use SR with google server which uses HMM method. The brief description of how speech is recognized is as follows. Firstly the speech is inputted, sound can be fluctuating set of signals which are recorded [2]. These signals depends on speaker how is his/her voice quality and hold on the language. The input data is divided into words and phrases, i.e. command is divided into several parts. Lastly comes the processing phase where accordingly system understands command and executes it.

Speech Recognition stands majorly on five pillars that are, feature extraction, acoustic models database which is built based on the training data, dictionary, language model and the speech recognition algorithm. The input data i.e. voice is first converted to digital signal and are sampled on time and amplitude axis. This digitalized signal is then processed. For processing the signal is divided into small intervals, which depends on the algorithm used. The generalized timestamp is

978-1-4673-8810-8/16/$31.00 ©2016 IEEE

20 ms. This division is based on the features of data as those features are compared with database element. Database element contains information of feature of the word found and according the command is created. The basic element can be a phoneme for continuous speech or word for isolated words recognition.

The dictionary or Database is used to connect the frequency model i.e. the spoken word with actual vocabulary word. The signal namely speech has its constraints as said speech should match the meaning of textual language brain created. The HMM uses word for modeling [3]. The output is a hidden probability function of the state which cannot be deterministically specified. States sequence is never a command that is SR system generally assumes that the signal is realization of message which is encoded as a sequence of symbols. Here symbols are the words sampled. To effect the reverse operation of recognizing the underlying symbol sequence given a spoken utterance, the continuous speech waveform is first converted to a sequence of equally spaced discrete parameter vectors. Vectors of speech characteristics consist mostly of MFCC (Mel Frequency Cepstral Coefficients), standardized by the European Telecommunications Standards Institute for speech recognition. The MFC can be easily created. The Fourier analysis is performed on sampled i.e. divided data then variable bandwidth triangular filters are placed along with the Mel frequency scale and energies are calculated by spectrum [1].

Lastly, the magnitude compression is applied and spectrum is de-correlated using the DCT and first coefficients represent the MFCC's. These vectors are called as observations which serve to future calculation. Finally all the states undergo same procedure one after another, all the states vectors are calculated and are checked with the dictionary and a command is created [4]. The system is trained through Iterative Baum-Welch procedure. The training is repeated until good accuracy is reached. The training should be accurate to an extent that if some minor changes are present the output can be generated with same efficiency. The decoding is done through Viterbi algorithm; it is according to output sequence which is defined by recursive relation created above.

A. Hidden Markov Models In Speech Recognition

An HMM algorithm is a random probability finite state automation defined by parameters. A HMM chain like depicted in the following figure 1. It has a set of N distinct states S1, S2, till SN, at regularly spaced discrete times i.e. the sampled signal. The time instants or the time at which signal is sampled are denoted with t = 1, 2, and so on. The actual state at time instance t is denoted as qt. The important feature because of which HMM is used is that this probabilistic model is Markov property which states that current state, future state and past states are independent to each other it means the states are calculated independently. Equation (1) and (2) defines the state transition probability as follows:

$$P\,[qt = Sj \mid qt\text{-}1 = Si,\ qt\text{-}2 = Sk\ldots] = P\,[qt = Sj \mid qt\text{-}1 = Si] \quad (1)$$

$$a_{ij} = P[qt = S_j \mid q_{t-1} = S_i], \quad 1 \le i, j \le N \quad aij \ge 0$$

$$\sum_{j=1}^{n} aij = 1 \quad (2)$$

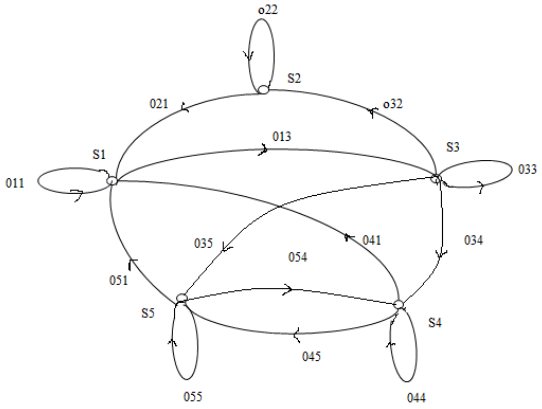

Fig. 1 Markov chain with states

The Markov chain is also called as observed chain as output process corresponds with the observed states only [5]. The HMM can be thought of as black box, where the sequence of output symbols produced over time is observable, but sequence of states visited over time is hidden view. That's why it is called a hidden markov model [6]. Simple HMM with two states and two output states A and B shows in below figure 2.

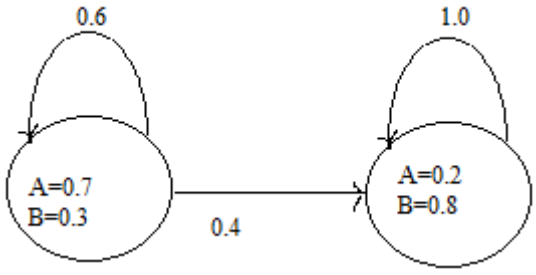

Fig. 2 Hidden Markov Model

The three basic problems which tend to arrive are as follows: First problem states that if observed sequence as O and model as λ, then how probability of the sequence with the model is efficiently computed. Second problem is that if observed sequence as O and model as λ, then how can inner state sequence Q explains the observations O. Lastly how to choose parameter to get the best and accurate output. Solutions to these are the Forward algorithm (to get the efficient output sequence); the Viterbi algorithm (to get the best observed state O) and the Baum-Welch algorithm (to choose the computing parameters) are used respectively.

III. IMPLEMENTATION

The system here is divided into small subsystems. In our application the models are as follows. Firstly user has an option to select whether to use voice as an input or select contacts manually. If user selects manual option then a service is called which access all the contacts in contact list present on the cell phone. If voice is selected then the google SR api is called and a dialog box appear which says that speck now and a mic type image is formed. Once the user is done speaking then api takes a few seconds to process the data and output is displayed on the sender's address block.

Here validation is applied, since the application sends message to number's therefore only numeric characters are allowed so that unnecessary data is trimmed out. Similar kind of work is done with message dialog. Here when the user presses message box user has an option to write the message manually or to import message via his/her voice. If manual option is chosen then keyboard appears and data can be inputted. If voice is chosen then again the google api is called and a dialog box appears saying speak now. After giving input, the inputted data can be seen in message box. Here no constraint is applied as message can be in numeric as well as alphabetical characters. Once both the fields are full in the end a button is present which says send message and once the user clicks it, message service is called and message can be sent to the inputted receiver.

Here the google api is always called, processing of api is as follows. The HMM rule is applied and then input is taken and sampled. A forward algorithm is applied to forward variable and partial observation is stored until time t with model λ. The forward variable is defined as follows in equation (3).

$$\alpha t(i) = P(O_1\ O_2 \ldots O_{t,\ qt} = S_i \mid \lambda) \ldots \quad (3)$$

Now we have to calculate highest probability from the given sequence which is to be stated by inner states of the sequence for that Viterbi algorithm is used and it helps in calculating the highest probability along a single path at time t. It is defined in equation (4) as follows.

$$\delta_{t(i)} = \max_{q_1, q_2 \ldots q_{t-1}} P[q_1, q_2 \ldots q_t = S_i, O_1, O_2 \ldots O_t | \lambda] \quad (4)$$

Lastly the parameters are chosen by the Baum- Welch algorithm and equation (5) is defined it as follows.

$$\xi t(i,j) = P(qt = Si, qt+1 = Sj \mid O, \lambda) \quad (5)$$

Once the output is processed it is displayed on the desired place.

IV. RESULT

Developed Speech recognizer system tested for a SMS sending application and found that it recognizes the speech to an accuracy of more than 90%. Enter phone number by speech or select contact from contact list. As user presses select contact here by selecting name of person it gives all phone numbers of that person in phone contact list box. Now it is possible to send sms to all numbers of same person on one click which results in reducing time of searching each number. When user presses the message box user has an option to write the message manually or to import the message via his/her voice. If manual option is chosen then the keyboard appears and the data can be inputted. If voice is chosen then again google api is called and a dialog box appears saying speak now as shown in figure 3.

978-1-4673-8810-8/16/$31.00 ©2016 IEEE

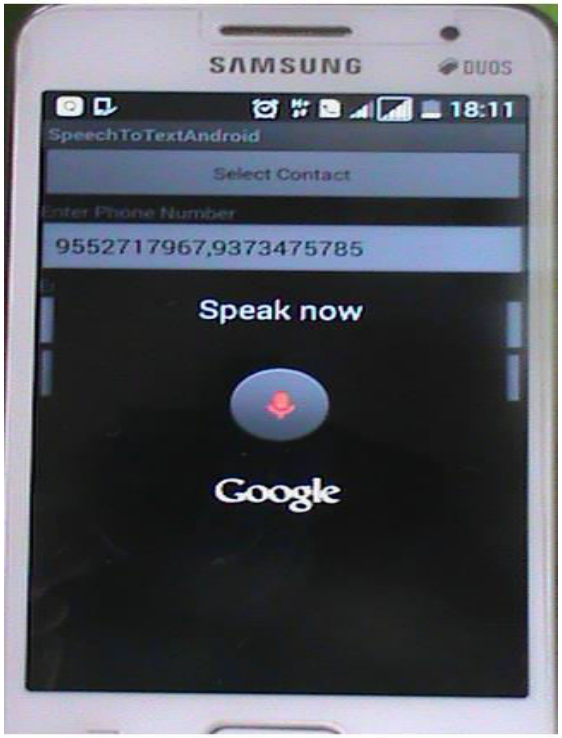

Fig. 3 Speech input

Some simple AI is also added to it by which checking validity of input speech for phone number. If the user is giving phone number as an input, but input speech is not a number, then the system gives recognized speech as well as informs user that this input is not a proper number as shown in given figure 4.

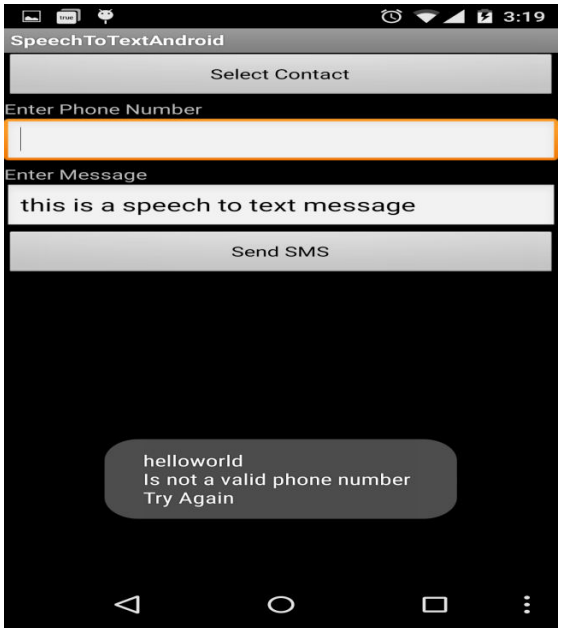

Fig. 4 Incorrect phone number input

This system tested for various speakers which had varying speech speed, amplitude and frequency. The results of this system are very good and recognized most of the speech inputs.

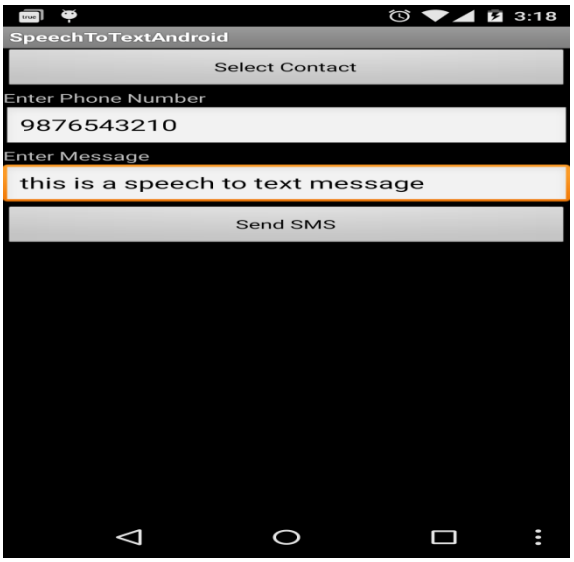

Fig. 5 Correct Inputs

V. CONCLUSION

An automatic speech recognizer studied and implemented on the android platform which gives much accuracy for both numeric and alpha numeric inputs. The accuracy of this system is about 90%, and delay for recognition is less than 100 ns.

We plan to implement this work for other languages as well as test them on the SMS sending application which is developed.

References

[1]. B. Raghavendhar Reddy, E. Mahender, "Speech to text conversion using android platform ", IJERA Feb-2013.

[2]. V.A.Mane, A.B.Patil," Comparison of LDM and HMM for an application of a speech", International Conference on Advances in Recent Technologies in Communication and Computing,978-0-7695-4201-0/10 $26.00 © 2010 IEEE, DOI 10.1109/ARTCom.2010.65-431

[3]. Brahim Patel, Dr. Y. SrinivasRao ,"Speech recognition using HMM with MFCC- an analysis using frequency spectral decomposion technique" , SIPIJ Dec 2010.

[4]. [1]Anjali Arora, [2]Anil Chopra," Query processing for hindi keywords searching using NLP", International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online) Vol. 3, Issue 2,Month: April - June 2015, pp: (981-984).

[5]. Patrick Gampp, "Hidden markov model basics".
[6]. J. Tebelskis,"Speech recognition using neural networks", Pittsburgh: School of Computer Science, Carnegie Mellon University, 1995.

978-1-4673-8810-8/16/$31.00 ©2016 IEEE