IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-4, Special Issue-7, Dec 2016 ISSN: 2347-2693

Speaker Independent Speech Based Telephony Service for Agro Service using Asterisk and Sphinx 3

Aniruddha Deka[1*], Manoj Kumar Deka[2]

Computer Science and Technology Department, Bodoland University, India
dekaaniruddha@gmail.com, manoj.deka1@gmail.com

Available online at: www.ijcseonline.org

***Abstract*— In this paper we describe the design of telephony Automatic Speech Recognition (ASR) System which can be used by low literate users. For this purpose we have deploy a spoken query system over Asterisk server using Bluetooth enabled mobile as client interface . Bodo speech data are collected by using the developed IVR module form BTAD area. All incoming calls are handled by Asterisk Server i.e. Computer telephony interface (CTI). The system asks to prompt 31 different words in Bodo which include Bodo digit, agricultural word and Yes/No. The collected speech is manually transcript and labeled for ASR system training. We discuss our experiences and fieldwork which includes design of the IVR using asterisk, Bodo ASR engine using sphinx and related study in the field of agricultural domain. We also evaluate low cost, rapidly deployable speech technologies for new languages as a means to improve equitable, affordable access to information technology (IT). The performance of a speech recognizer built using Bodo language is evaluated on the basis of word error rate and we 77.24 % word accuracy in training mode and 72.12 % in testing mode.**

***Keywords*— Signal Processing, Telephony Application, Asterisk server, Interactive Voice Response**

I. INTRODUCTION

Information is a basic necessity of everyday life. From a variety of sources information can be obtained or retrieved. Farmers are a particular group information needs is very specific. Due to the lack of information retrieval system farmer community in rural areas they prefer other colleague or fellow farmers for required information following by newspapers and Government office. In their day to day agricultural concerns, they take decisions to solve problems. However sometimes they take wrong decisions and make mistakes due to the unavailable, irrelevant and insufficient information. For these reasons despite having fertile land, huge and hard-working labor force, and other resources, farmers cannot make proper contribution to the national development of the country. [1] The lack of knowledge of information needs of a particular community is a major obstacle in the design of need-based information services that can provide more relevant information to its users. Understanding farmers' information needs and seeking behavior is an important first step in designing focused, need based, and user-oriented information infrastructure in the agricultural sector. [2] [3] [4]

During survey we found the following problems for accessing agricultural information by the farmers [2]:

- Lack of timely access to the required information
- Low level education
- Language barrier
- No native language based system
- Lack of awareness about where to get required information
- Electricity load-shedding and bad timings of programmes
- Infrequent visit of extension staff in the village

II. MOTIVATION

Modern human life is totally dependent on portable technology like mobile, PDAs, GPRS etc. Now a day's Voice controlled public purpose emergency information retrieval services are also demanding. For example Weather forecasting, Traffic reporting, Travel enquiry system, Health informatics etc. accessible via mobiles to fulfill urgent and on the spot requirements [5].But for real time deployment of the system there is a need of speech based interface module with pretty good recognition or some DTMF based application. Again human voice or speech is the fastest communication form in our daily busy schedule that further extends the usability of such voice enabled mobile applications in emergency situations. In such a scenario, speech-centric user interface on smart hand-held devices is currently foreseen to be a desirable interaction paradigm where Automatic Speech Recognition (ASR) is the only available enabling technology [5]. Interactive Voice Response (IVR) systems provide a simple yet efficient way for retrieving information from computers in speech form through telephones.

A practical IVR system should be designed in such a way that it should be capable of handling real time telephony hazards like channel drop, clipping, speech truncation etc. It should also provide robust performance considering issues like Speaker-Variability, Pronunciation /Accent Variability, Channel Variability, Handset Variability, Different Background Noise etc. Considering the above said requirements, telephonic ASR is being designed in such a way that, to-some-extent it can meet the above mentioned capabilities.

In the present study, speech data are mainly collected from all the geographical regions where native Bodo language speaking population is considerably high. The collected speech data is then verified and used for ASR training. The reason behind choosing a large geographical area for data collection is to cope up with the problem of speaker variability, accentual variability. Additionally, various issues regarding the telephonic channel such as channel drop or packet lost during transmission, handset variability, service provider variability, various types of background noise such as cross-talk, vehicle noise etc. have been observed, analyzed and estimated efficiently from the collected speech data and modeling of those can improve ASR performance. These issues will not only help us to improve the system performance effectively, but also provide us very good research motivation on other telephonic applications.

III. RELATED AGRO SERVICES

In our literature survey we found that there is lots of system developed in many countries which gives the information of agricultural commodity to farmers as well as users. Some of the application is web based, some are mobile based and some IVR based. Some of the application is design in their local language. But we don't find any application that build in tonal based language. Few of them have been shown to work efficiently and have resulted in helping the farmers. Some of the systems that were developed in order to facilitate the farmers are given below:

SAPA Mobile is a mobile-based farm advisory information system. Application stores the need information of different market, so farmers can directly linked or search the markets and sell their crop in good price which give them more profit. *Nokia life tools* is a SMS based service tools which is build for agriculture, weather, education and entertainment information. User can able to get the information in local languages. It gives agricultural information of for higher productivity and higher earnings. *AgriFone* is a non commercial mobile based solution which addresses the needs of farmers, agricultural workers, agribusinesses and input suppliers. This application gives convenient and easy to use system which can be accessed with cheap mobile phones. Unique feature of AgriFone is that farmers can exchange text, voice and images in one-to-one, one-to-many and peer-to-peer basis amongst mobile subscribers [4]. *MKrishi* is a mobile based proprietary solution designed by TCS. This system allows farmers to send their agricultural query to agricultural experts and received advice in their own local language. They design this system by integration of different technology like sensor, DMA modem, CDMA network, solar power, handset with camera, and an engine to assist to displaying mobile screens in Indian languages. *0700 Interactive voice response system (IVR) [4]* service is launched only to facilitate farmer community but was not able to generate sizeable revenue therefore this service is currently discontinued. *ESOKO* is mobile based system developed for sending SMS to farmers to provide them agricultural information. The different services provide by ESOKO are live market feeds, direct SMS marketing, Scout pooling, profiling and marketing.

The *E-Arik* project in Arunachal Pradesh gives information about the crop cultivation and other agricultural practices. This project gives information about the specific information on government schemes such as farmer welfare programmers. It also gives weather forecasts and day to day market information. M-Farm is a mobile based application which helps farmers of Kenya to get market information and improve their agriculture productivity. This application gives different functionality to farmers like price information, group selling, and group buying and customer relationship management. Reuters Market Light offers customized information on commodity prices, local news and weather updates .It works across mobile network operators. *E-Choupal* gives crop management advisory services to individual farmers by integrating mobile phones into the digital and physical network of echoupal. [4]

IV. PROPOSED WORK

Our proposed system consists of Asterisk server, SDS system and database. Asterisk is open source software on Linux/Unix platform that enables the computer to be used as a telephone network server [4]. The Asterisk server in turn consists of computer telephone interface (CTI) card and IVR. The Asterisk [7] PBX server is implemented on Ubuntu 10.04 Linux. There is a series of configuration files which controls Asterisk. It uses several directories on a Linux system to manage its various aspects, such as voicemail recordings, voice prompts, and configuration files. The necessary directories are:

- /etc/asterisk - Contains all of asterisk configuration files and logic information.
- /usr/lib/asterisk/modules - Contains all of asterisk's loadable modules, operating asterisk functionality.
- /var/lib/asterisk/sounds - Contains all of asterisk's sound files for playback and pre-loaded applications.
- /var/lib/asterisk/agi-bin - Contains all of asterisk's

AGI scripts and AGI logic.

Asterisk Channel Driver to allow Bluetooth Cell/Mobile Phones to be used as FXO devices and Bluetooth Headsets as FXS devices. Asterisk is connected with network by Using Bluetooth enabled mobile network. In Asterisk the Chan-mobile supports FXS and FXO station interfaces for connecting wireless lines through a PC.T1 card that allows a digital trunk to be connected to Asterisk. All of these cards allow making calls directly on the Public Switch Telephone Network without have to use a VOIP phone service provider. [7] The zttool program can be used to check the status of installed hardware in Asterisk. The configuration information is stored in the following files:

- /etc/zaptel.conf: Configuration of your hardwareInterfaces
- /etc/asterisk/zapata.conf: Asterisk configuration touse your hardware interfaces

The "extensions.conf" is the configuration file of Asterisk. It contains the "dial plan" of Asterisk, the master plan of control or execution flow for all of its operations. Dialplan controls how incoming and outgoing calls are handled and routed. This file configures the behavior and control of all connections through your PBX. The content of "extensions.conf" is organized in sections [10][11][12]. The settings sections are general and global and the names of contexts are entirely defined by the system administrator. After registering, an appropriate extension number is dialed and a call is received at the server end. Asterisk server, upon receiving the call, executes the dial plan written specific to the called extension number and plays an audio file requiring the client to speak any input as prompted by the server which is further recorded in a wav file format for 3 sec (specific to application) and stored in a fixed location. In the next step of dial plan, a shell script is invoked which externally provides output to be further used by Asterisk server as input to process user defined logic implementation. Finally the stored file is used as input for recognition in offline recognition mode, given in the shell script .The recognized word is stored in a text file (i.e. output.txt). Fig1 gives the detailed call flow structure of the proposed system.

To convert the call flow into dialogue manger we have used PHP-AGI script which is integrated with extention.config file. Once the call landed to extention.config file the script will activate and start working according to flow.

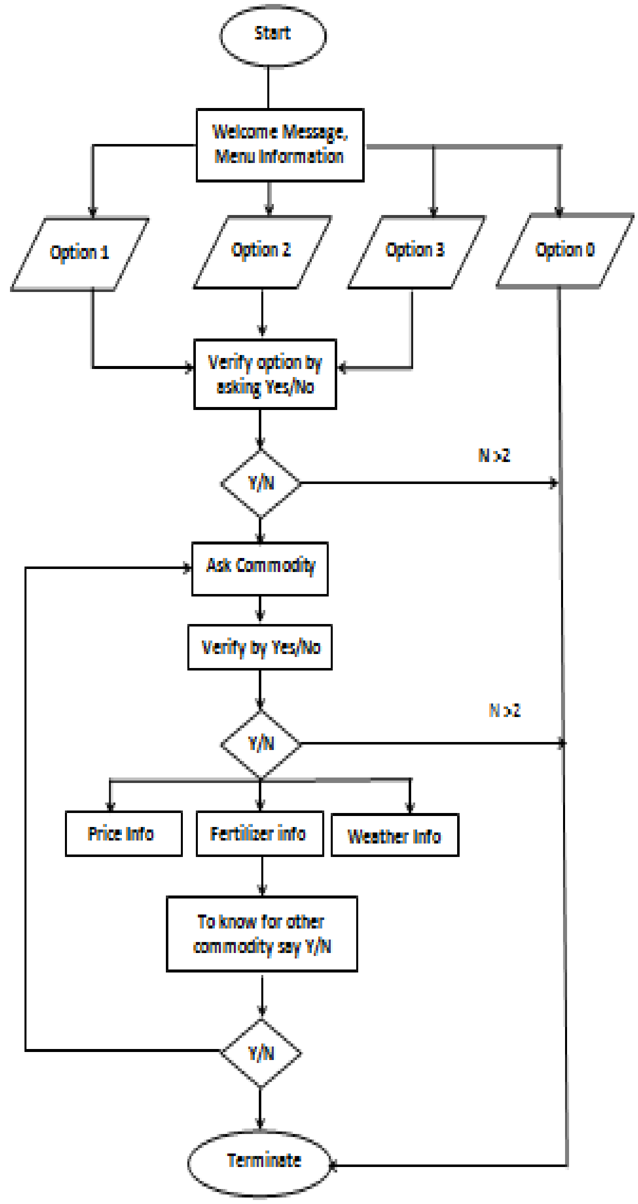

Fig: 1 Call flow of the proposed system

Following steps gives the complete IVRS flow of our proposed agro information system.
System: Welcome to Agricultural information system. To know the price of a commodity Say 1, to know about the fertilizer of the commodity say 2, to know the weather information about the commodity please say 3 and to end up the call please say Zero.
User: one
System: You have chosen one. To confirm it say yes or to deney say no.
User: yes
System: What is the commodity name for which you want to know the price?

User: Potato
System: You have said Potato. To confirm it say yes or to deny say no.
User: Yes.
System: The whole sale price is 1000 per quintal and the retail price is 12 per kg.
System: To know the price of another commodity please says yes or to go to main menu say no.
User: No
System: Welcome to Agricultural information system. To know the price of a commodity Say 1, to know about the fertilizer of the commodity say 2, to know the weather information about the commodity please say 3 and to end up the call please say Zero.
User: Two
System: You have chosen two. To confirm it say yes or to deney say no.
User: yes
System: What is the commodity name for which you want to know fertilizer information?
User: Potato
System: You have said Potato. To confirm it say yes or to deny say no.
User: Yes.
System: ………………………………………………….
System: To know the fertilizer information of another commodity please says yes or to go to main menu say no.
User: No
System: Welcome to Agricultural information system. To know the price of a commodity Say 1, to know about the fertilizer of the commodity say 2, to know the weather information about the commodity please say 3 and to end up the call please say Zero.
User: three
System: You have chosen three. To confirm it say yes or to deney say no.
User: yes
System: What is the commodity name for which you want to know Weather information?
User: Potato
System: You have said Potato. To confirm it say yes or to deny say no.
User: Yes.
System: ………………………………………………….
System: To know the weather information of another commodity please says yes or to go to main menu say no.
User: No
System: Welcome to Agricultural information system. To know the price of a commodity Say 1, to know about the fertilizer of the commodity say 2, to know the weather information about the commodity please say 3 and to end up the call please say Zero.
User: Zero
System: Thank You.

V. DATA COLLECTION AND PREPARATION

As per requirement of our system we mainly required three types of data, commodity name, digit and yes/no. Our speech corpus consist of 31(25+4+2) unique Bodo words, out of which 25 commodity name, 4 digit (0, 1, 2, 3) and Yes/No word. There for we identify 100 different Bodo speaking people for different location and collect data in quite silent environment. We take 70 males speaker and 30 female speaker of different age group for data recording purpose. The Sphinx toolkit does not require phoneme duration information for the training sentences, the (differences in) timing in the pronunciation of the training sentences is not important. The sphinx toolkit learns to recognize the words through fitting the word transcriptions on the training set. These transcriptions are used for all realizations of the same sentence, even though there might be variation between speakers relative to the transcription. The speech data was digitized at 16 bits/sample at a sampling rate of 16 kHz. The speakers were advised to read at their comfortable speaking rate. The speakers were given a list with agricultural commodity words, Bodo digit and yes/no sample which they had to read aloud. Finally we have collected 7171 words from 100 speakers which we are going to use speech recognizer. The data is recorded with the help of a unidirectional microphone using a recording tool wave surfer in .wav format.

For training acoustic models is necessary a set of feature files computed from the audio training data, one each for every recording in the training corpus. Each recording is transformed into a sequence of feature vectors consisting of the Mel-Frequency Cepstral Coefficients (MFCCs). Apart from the Wav files Sphinx 3 requires transcription File, Pronunciation dictionary, Filler dictionary for training of the system. So these files are created according to the required format.

VI. FEATURE EXTRACTION,TRAINING AND TESTING

A speech signal can be modeled as a series of piecewise stationary signals, which can be decoded to a set of one or more symbols. The continuous speech waveform is first preprocessed and converted into a sequence of discrete parameter vectors, called speech vectors. The role of the speech recognition system is to create a mapping between the speech vectors and the corresponding symbols. This system was trained using 5171 files of commodity and digit words collected from 100 speakers. The various parameters that are used for feature extraction are summarized in Table 1

Parameter	Default Value	Modified value
Pre emphasis coefficient	0.97	0.97
No of filters in filter bank	40	31

Lower cut –off freq	130 Hz	130
Upper cut-off freq	6800 Hz	3500
Sampling rate	8000 Hz	16000
Base feature Dim	13	13
Window length	0.00256 sec	0.00256 sec
Frame rate	100 per sec	100 per sec

Table 1: MFCC feature extraction parameters

The speech recognizer is is developed using the open-source speech recognition toolkit Sphinx-3. The diagrammatic representation of the training and decoding modules are shown in Figure 4 and Figure 5 respectively. The tools SphinxTrain-1.0 and Sphinx3.0.8 are used for training and decoding purposes respectively. MFCC features of 39 dimensions, comprising of 13 base features with its first and second derivatives, are used for speech parameterization. During training time system shinxtrain verify the following things.

a. Checking to see if the dictionary and filler dictionary agrees with the phonelist file.
b. Checking to make sure there are not duplicate entries in the dictionary
c. Check general format; utterance length (must be positive);
d. Checking number of lines in the transcript should match lines in control file (*.fileids)
e. Determine amount of training data, see if n_tied_states seems reasonable.
f. Checking that all the words in the transcript are in the dictionary
g. Checking that all the phones in the transcript are in the phonelist, and all phones in the phonelist appear at least once

VII. RESULTS & DISCUSSION

In our project the total vocabulary consist of 31 words. We have collected a total 6171 utterance words taken from different Bodo speaker. Including these words we have taken 5171 utterance words for training, which are spoken by 100 different users and took them as a trainee in training phase by recognition toolkit. Along with the keyword models one separate filler model is also created to model ground noise and false starts by users. A sphinx 3 toolkit is used to construct Tied state Triphone models. Mixture incrementing is done in successive iterations with models testing at each iteration for optimum results. The recognizer is then integrated with Asterisk server by calling Asterisk gateway interface (AGI) program using shell script. The implemented system is then tested by giving voice input through the mobile phone. The following tables give the results of different experiment that we perform.

Mode	Word Accuracy	No. of deletion	No. of Substitution	No. of Insertion
Training	77.24	3	2	0
Testing	72.12	7	8	2

Table 2: Word Accuracy with training and testing data

Decoder	Sennon 1000	1500	2000	2500	Total train files
Commodity	70.23	71.34	72.38	72.69	5171
Bodo-Digit	87.3	84.6	86.3	86.2	1500
Yes/No	93.1	93.2	89.3	93.4	350

Table 3: accuracies with diff no .of tied states and 16 GMMs/state

Decoder	Sennon				Total test files
	1000	1500	2000	2500	
Commodity	69.5	62.8	72.8	72.6	100
Bodo-Digit	82.1	80.7	80.2	82.2	500
Yes/No	89.2	88.3	91.3	92.4	50

Table 3: Accuracies with diff no. of tied states and 32 GMM

VIII. CONCLUSION

In this work, the speech recognition system in Bodo along with the Bodo language model has been developed and tested. The trainer and the decoder configuration files have several parameters; these parameters can be tested to improve the efficiency of the speech recognition system. Besides, other than Bodo, this speech recognition system can be further extended to include other regional languages as well, by developing their corresponding language dictionaries. This will add to the Development of Multimodal User Interface (DMU). Apart from the Bodo recognizer the sphinx and asterisk based IVR design architecture also described in this paper.

REFERENCES

[1] Nitin Bhagachand Bachhav," Information Needs of the Rural Farmers : A Study from Maharashtra, India: A Survey" , in Library Philosophy and Practice (LPP) (ISSN 1522-0222), 2012

[2] Muhammad Asif Naveed and Mumtaz A. Anwar , “Agricultural Information Needs of Rural Pakistani Farmers ”, in Malaysian Journal of Library & Information Science, Vol. 18, no. 3, 2013: 13-23.

[3] Assam Small Farmers Agri-Business Consortium, http:// assamagribusiness.nic.in.

[4] Agricultural Marketing Information Network - AGMARKNET,http://agmarknet.nic.in.

[5] Joyanta Basu, Milton S. Bepari, Rajib Roy, Soma Khan, " Telephony Speech Recognition System: Challenges" in National Conference on Communication Technologies & its impact on Next Generation Computing CTNGC 2012, Proceedings published by International Journal of Computer Applications® (IJCA).

[6] https://mahtabrasheed.wordpress.com/2012/11/07/mobile-phones-based-agricultural-information-systems, accessed in Septembor, 2016

[7] Asterisk.org, "Features of Asterisk PBX", http://www.asterisk.org/features, accessed in Septembor, 2016

[8] http://cmusphinx.sourceforge.net, accessed in Septembor,2016

[9] Gautam Varma Mantena, S. Rajendran, B. Rambabu, Suryakanth V. Gangashetty, B. Yegnanarayana, Kishore Prahallad, "A speech-based conversation system for accessing agriculture commodity prices in Indian languages", IEEE Explore , ISBN: 978-1-4577-0999-9, 2011

[10] S Shahnawazuddin, Deepak Thotappa, B D Sarma, A Deka, S R M Prasanna and R Sinha,"Assamese Spoken Query System to Access the Price of Agricultural Commodities", NCC 2013, IITDelhi.

[11] Asterisk open source communications, http://www.asterisk.org/home.

[12] S. Goel and M. Bhattacharya, “Speech based dialog query system over asterisk pbx server,” in 2nd International Conference on Signal Processing Signal Processing Systems (ICSPS), Dalian, Jul 2010.

AUTHORS PROFILE

Aniruddha Deka, received the B.E degree in Computer Science from North Eastern Hill University from Shillong in the year 2006 and M.Tech degree in IT from Gauhati University, Guwahati in 2013.He worked as an Assistant Project Engineer in IITGuwahati from the year 2007 to 2012, mainly focusing on development of speech based application in Assamese Language. He is currently pursuing the Ph.D. Degree in Computer Science and Technology at Bodoland University, Assam. He has 19 national/ international journal / confernce papers and attended 18 numbers of workshop/conference attended till date. His research interest includes speech processing, Database management system, Web programming.

Dr. Manoj Kumar Deka, received the M.Sc. degree in Instrumentation in the year 2007 and Ph.D. degree in Signal Processing in the year 2011 from Gauhati University, Assam. He achieved INSPIRE Fellowship sponsored by Department of Science & Technology (DST), Govt. of India, New Delhi during his Ph.D. work. Presently he is working as Assistant Professor in the Department of Computer Science & Technology, Bodoland University, Assam. He has already published 19 numbers of international and national papers. He presented large numbers of research papers in International and National Regional conferences and seminars. He participated 5 numbers of National and International workshop till date. Five numbers of research scholars are now pursuing their Ph.D. research work under his guidance.His area of interest includes Signal Processing, Embedded system, Artificial Intelligence.

