1

OVERVIEW OF THE THESIS

CONTENTS

1.1 INTRODUCTION …….....2

1.2 OBJECTIVE OF THE THESIS ……3

1.3 BODO LANGUAGE ……4

1.4 MOTIVATION FOR CHOOSING BODO LANGUAGE ……..5

1.5 SOCIAL PROBLEM STATEMENT ……5

1.6 RESEARCH PROBLEM STATEMENT ……..……6

1.7 INITIATIVE IN AGRICULTURAL DOMAIN ……...7

1.7.1 MOBILE-BASED INITIATIVE…….……...8

1.7.2 INTERNET-BASED INITIATIVE ……..…….8

1.7.3 COMMUNITY RADIO STATION ……..…….....8

1.7.4 CALL CENTRE/ IVRS ……9

1.8 SURVEY ON AGRO SERVICES ……...……9

1.9 COMMODITY PRICE DECIMATION SYSTEMS…..…….....11

1.9.1 PROBLEM WITH THESE SYSTEMS ….....11

1.10 CONTRIBUTIONS OF THE THESIS ……12

1.11 ORGANIZATION OF THE THESIS ……12

1.1 INTRODUCTION

Information technology and its applications are essential part of our day to day life [1]. We use computer and computer-based application like ATM, paying bills, online purchasing, flight booking etc. for our regular routine works. Most of these applications contain graphical user interface. These applications takes input through keyboard or mouse and finally get the response. Cost and current methods of input [2] i.e. mouse, keyboard, styles, mobile phone keypads create a barrier to information technology for the majority of illiterate people to take the advantage of it. General people who are not habitual of the recent technology are not able to use these kinds of system. This inability put some question to the researcher that:

a. Could we somehow achieve human-like performance and flexibility with computer interfaces?

b. Wouldn't it make sense to have systems that could be used by having human-human like conversation, that is, the most natural way for a human to interact with somebody?

One of the solutions of these questions is spoken dialogue system. Spoken dialogue systems try to achieve this conversational gap between the user and the system. It minimizes the gap for literate/illiterate and visually challenged users to access information. Speech based user interfaces are often suggested as a means to overcome current barriers of cost and literacy. Speech-based UIs are cheaper than display-based UI solutions and more accessible to illiterate or semi-literate users than text-based UI solutions. However successful speech-based UIs in developing regions are rare due to the unique challenges and constraints from data collection to deployment including lack of training data, multilingual, dialectal and cultural diversity and lack of resources. In addition, power, connectivity and limited infrastructure are all significant obstacles in developing regions. UI design also requires a familiarity with the cultural and economic context of the user. As a whole, people who have never learned to read or write are poorly understood by researchers and developers of technology. For the deployment of such kind of system, there is need of decoder module with pretty good recognition accuracy. Apart from the recognition, the system 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 [3], handset variability, different background Noise etc.

1.2 OBJECTIVE OF THE THESIS

The first objective of this thesis is to design and develop a speaker independent spoken dialogue system that will provide relevant information in an agricultural domain to Bodo people, in a convenient manner. At each point, a directed question is asked, and the user would reply with an appropriate response from a small set of words. The system would use a set of directed questions, so as to elicit an appropriate response from the user. The directed questions will be organized in a tree-like structure of minimal depth, so that at any point only a small set of words needs to be recognized, resulting in an effective system, that achieves high transaction success. The performance of the system would be measured by the percentage of times it successfully provides the relevant information without going to a human operator. But developing such kind of system for research purpose with low cost is not an easy task. So building a low cost spoken dialogue system for research purpose is the second objective in my research. Channel drop, clipping, speech truncation, speaker-variability, pronunciation /accent variability, channel variability handset variability, different background noise are also certain issues for building spoken dialogue system. In this work, speech data are mainly collected from all the geographical regions where native Bodo language speaking population is considerably high. The reason behind choosing a large geographical area for data collection is to cope up with the problem of speaker variability, accentual variability, etc. The analysis of all those issue is the third objective of our work.

During the system building phase, domain-specific data had been collected and the segmentation, pitch & intensity techniques for word boundary detection have been analyzed in Bodo speech. The speech-interface will also be designed to take into account speaker idiosyncrasies, as well as errors of the ASR system. The recognition performance is evaluated with different speech features and adaptation techniques.

1.3 BODO LANGUAGE

Bodo (suggested as boron) is a language of north east India spoken through the Bodo tribe of people that is a prominent ethnic organization inhabiting Assam, Bhutan, Nepal and Bangladesh. It falls under the Assam-Burmese group of languages. It branched off from the Tibetan and Burmese family of languages in the past. It is one of the twenty-two recognized languages by the Eighth Schedule of the Indian Constitution and is among the official languages of the Indian State of Assam [4].Due to its tonal behavior there are plenty of scopes for research. The Bodo language has different special characteristics such as:

- **a.** It has intonation pattern, juncture and two types of tones.
- **b.** The words in Bodo are highly monosyllabic.
- **c.** It has agglutinative features.

This language has a total of 22 phonemes: 6 vowels and 16 consonants.

- **a.** Vowels: अ, आ, इ, उ, ए, औ
- **b.** Consonants: ख, ग, ङ, ज, थ, द, न, फ, ब, म, र, ल, स, ह
- **c.** Semi Vowels: य, व

Bodo is a tonal language. Tones refer to the distinctive pitch level of a syllable. In many languages the tone carried by the word is very essential for the meaning of the word. If the meaning of a word is changed by the pitch of word, that language is called a tone language. The pitch can change nuances as well as core meaning of the language. The different tones are produced by the different pitch level of the language. For example

Anjalua_____ bungdungmon

Anjalu _____ said

The acoustic result of the speed of the vibration of the vocal cord in the utterance of the voiced part of the sound produced pitch. The rapid vibration of vocal cord produces high-pitched sound and slow vibration produces low-pitch sound. The tones may fluctuate due to pitch contour movement and thus raising and falling tones are produced [4]. All languages have pitch variation. The function of this pitch is different from language to language. The pitch difference used in this way is called tones. Tone may be on a single level of pitch, called level tone or may fluctuate and thus produce contour type of tones. As a result of the fluctuation, the level of tone

may change and produce different categories of tones. If the pitch level rises during the articulation of the sound it is called rising tone. If the pitch level falls, the tone is called falling tone. There may be fluctuation in the middle to produce the tones rising-falling and falling-rising. Based on the pitch movement from the starting position, the tones may also be classified as mid-level, high-level and low-level due to their level-wise movement or they may be mid-rising, mid-falling, high-rising, high-falling, low-rising and low-falling due to their fluctuation from the starting position. No standard or no uniform spelling system, splitting of words, joined sentence, less vocabulary, dialect variation, grammatically incorrect sentence, hyphenated words are some of the challenges in building Bodo corpus.

1.4 MOTIVATION FOR CHOOSING BODO LANGUAGE

In 21st century, it is one of the strongest appeals to make communication without language barrier, and moving up the knowledge chain. All the technocrats, scientists, literary bodies' government should give enough importance to language research to support local and regional languages. Choosing Bodo as the language for recognition comes from local relevance. The following are some of the motivations of the present study:

a. Very less study has been done in speech processing point of view in Bodo language.

b. To analyze the different speech feature and properties to detect the word boundary and to create a scientific corpus for research purpose.

1.5 SOCIAL PROBLEM STATEMENT

Due to the lack of information retrieval system, farmer communities in rural areas prefer other colleague or fellow farmers for required information, where they may follow newspapers and Government offices for the information. In their day to day agricultural concerns, they take decisions to solve problems. However, sometimes they take wrong decisions and make mistakes due to unavailability, irrelevance, and insufficiency of information. For these reasons, despite having fertile land, huge and hard-working labor force, and other resources, farmers cannot make a proper

contribution to the overall development of the country. 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.

In literature survey, we found that there is over 80% of the population directly engaged in agriculture in the Bodoland Territorial Council (BTC) area, where most of the people are belonging to the Bodo Tribe and they cultivate different commodities. The Ministry of Agriculture, Government of India, maintain a site named AGMARKNET which lists the price of commonly grown/traded agricultural commodities for all the major states of India and is updated frequently. In BTC area the majority of farmers are semi-literate or not computer savvy. So it is very difficult for them to browse the AGMARKNET website for required information. Another problem with the web-based system is that, it is not in native language. Due to the lack of timely access to the required information, low level of education, language barrier, no native language based system, lack of awareness about where to get required information, electricity load-shedding and bad timings of programs and infrequent visit of extension staff in the village, the rural farmers face problems for accessing the web-based applications [5-7]. For this problem, we try to investigate other solutions, through which we can potentially solve the issue. We found that some of the SMS, DTMF, IVR, and web-based solutions are there to solve these problems, but those are not cost effective and they are not designed for Bodo language. For this issue, we design a cost-effective agricultural spoken dialogue system, in Bodo language, so that Bodo users are able to communicate naturally and efficiently.

1.6 RESEARCH PROBLEM STATEMENT

The proposed work focus on the following issues

a. As boundary detection is a trivial task and one algorithm cannot be applied to another language due the different linguistic features hence there is a necessity to analyze different speech features in Bodo language**.**

b. In speech recognition system the feature extraction and the classifier techniques has the important role to increase the accuracy of the decoder module. The performance of the decoder always varies in controlled environment and uncontrolled environment. As our proposed work is based on real time scenario hence different speech features are analyzed to see the performance of the system in real time environment.

c. The developed system should handle speech from any arbitrary Bodo speaker, i.e., it would be a speaker-independent ASR system.

d. Variability in mobile handsets due to differences in spectral characteristics. Different handsets (mobile phone manufactures/models), proprietary compression techniques used by mobile service providers add to this problem.

e. Various kinds of noise, so that it is robust to real-world applications. The application is likely to be used by people in a variety of environments; the ambient noise decreases the SNR, resulting in decreased recognition accuracy.

f. Whereas a first-time user may need somewhat detailed instructions on the use of the system, an experienced user may want to save time by speaking even before the system completes its instruction. The current system will permit such barge-in. Speech disfluencies such as “uh”; “uhm” etc. will be handled by training acoustic models for common speech disfluencies, a strategy similar to that with respect to background noise.

1.7 INITIATIVES IN AGRICULTURAL DOMAIN

Information is a basic necessity of everyday life. Now-a-days there are various kinds of sources to obtained and retrieve information for the people who are IT savvy. But farmers need some special kind of information retrieval system for cultivation domain. Due to the lack of such systems, farmer community in rural areas is not able to get the updated agricultural information. Although government has taken various initiatives those are not sufficient for the farmers due to some limitations in those systems. Some of those initiatives and their limitations are explained in the following section:

1.7.1 MOBILE-BASED INITIATIVE

Normally in this type of initiatives information are provided in text mode only. But due to illiteracy and language barrier farmers are more comfortable with voice mode communication. There should be a need of new system which supports both text and voice. Also system should have some facility for registration of farm and personal details of the farmer so that personalized care can be provided to each farm. How it is found that in most of mobile based system information are provided in one way only. But there is of two way communication so that farmers can properly convey their problem in preferred mode and time. Only information dissemination through the mobile is not sufficient. To make useful of this kind of system, crop experts should follow up the farmer's queries on a regular basis.

1.7.2 INTERNET-BASED INITIATIVE

Information access through internet by farmers in rural areas is a major challenge for information decimation point of view. Regular orientation and training is required on the usage of internet technology. Most of the websites are not updated regularly. So information provided by these websites is also not generic. There for with the help of multilingual features such kind of websites should be make in regional language and information should be provide more specific manner. Such kind of system should be implemented in rural areas instead of urban areas with a facility of internet based kiosks. A provision should be given for farmers to consult directly with experts as it would be more beneficial for them and help them to get immediate solutions for their critical problems [8].

1.7.3 COMMUNITY RADIO STATION

"Community radio is a type of radio service that caters to the interests of a certain area, broadcasting content that is popular to a local audience but which may often be overlooked by commercial or mass-media broadcasters (UNESCO 2002".These kind of systems are the vehicle or system for delivering useful information to farmers and assisting those farmers to develop requisite knowledge, skills and attitudes to make use of this information or technology effectively. But they

provide one way communication due to which farmers could not get immediate solutions to their specific problems. The other challenges of CRS includes competition with other commercial stations, limited time of broadcasting, repetition of programs, and the sustainability of community radio stations. To overcome these issues coordination is required with local government agencies, agriculture universities and research centers involved in agriculture development to provide need specific programs on agriculture, to the beneficiaries. More promotional and awareness activities are also required to increase awareness followed by the participation of progressive farmers [8].

1.7.4 CALL CENTRE/IVRS

In this kind of initiatives due to the lack of farm and farming details mostly general information is provided instead of area specific information. There is not any feedback mechanism to ensure the quality and effectiveness of responses that farmers get from experts. The incorporation of image, video and text mode is required to increase the popularity of these initiatives. This will help farmers and experts to communicate and understand the exact problem in specific cases. Proper monitoring and field visit by the AGRI expert is required to proper utilization of this system.

1.8 SURVEY ON 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 are IVR based. Some of the applications are designed in their local language. But we don’t find any application that builds in tone 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 [8-9]:

SAPA Mobile [10] is a mobile-based farm advisory information system. This application stores the need information of different market, so farmers can directly link or search the markets and sell their crop at a good price which gives them more profit. Nokia life tools is an SMS based service which is built for agriculture, weather,

education, and entertainment information. Users are able to get the information in local languages. It gives agricultural information for higher productivity and higher earnings. AgriFone [9] 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 a system which can be accessed with cheap mobile phones. The 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.

Kristi [11] is a mobile-based proprietary solution designed by TCS. This system allows farmers to send their agricultural query to agricultural experts and receive advice in their own local language. They designed this system by integration of different technologies like sensor, DMA modem, CDMA network, and solar power, handset with the camera, and an engine to assist in displaying mobile screens in Indian languages. 0700 Interactive voice response system (IVR) service is launched only to facilitate farmer community but was not able to generate sizeable revenue; therefore, this service is currently discontinued. ESOKO [9] is a mobile-based system developed for sending SMS to farmers to provide them agricultural information. The different services provided by ESOKO are live market feeds, direct SMS marketing, Scout pooling, profiling, and marketing.

The E-Arik [12] work in Arunachal Pradesh gives information about the crop cultivation and other agricultural practices. This work gives information about the specific information on government schemes such as farmer welfare programs. 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.

1.9 COMMODITY PRICE DECIMATION SYSTEMS

Over the last few years, there have been many telecom operators who have tried to make the price of the agricultural commodities easily available to the farmers. Some of these include:

a. Shyam Telecom and NCDEX: Shyam Telecom has combined with the National Commodities Exchange (NCDEX) to provide the following service in Rajasthan. By paying a small fee, farmers can access commodity prices from the local PCO operated by Shyam Telecom. This saves them a trip to the local mandi, besides losses from selling gullibly below market prices. The idea is to take price information deep into the hinterland, to a target audience that has little likelihood of accessing it otherwise. According to this scheme: PCOs will carry the prices as a ticker or a paid service through the Net where a farmer can come to pay an amount and access the prices.

b. Reliance and NCDEX: Reliance Infocom plans to use their R-World service to provide access to commodity prices. This is again done in collaboration with NCDEX. According to Reliance a sizable farmer community that have not had access to real-time information on PCs or the Internet will benefit from this new initiative on Reliance mobiles.

1.9.1 PROBLEM WITH THESE SERVICES

In both these services, there still exist major drawbacks:

a. In the Shyam Telecom scheme the farmer has to still have access to Internet and he has to go to a PCO to get the information.

b. In the case of Reliance scheme, the farmer should have capability to use R-world, and also should have reading/writing skills to be able to get the information.

This thesis aims to overcome these limitations and allow the farmers to get the information by just speaking into the mobile phone.

1.10 CONTRIBUTIONS OF THE THESIS

Following are the points of contribution of the thesis

- **a.** Analyze the pitch, intensity, ZCR, STE, Fo, hamming window and energy features in Bodo speech for voiced –unvoiced classification and boundary detection.
- **b.** ASR engines: Sub-word Hidden Markov Model (HMM) based ASR engines for recognition of agricultural commodity names Bodo language.
- **c.** Telephony data acquisition setup: Bluetooth based IVR for recording task-specific speech data over analog mobile channels and exchange of data with ASR system.
- **d.** Spoken language corpora: Multi-speaker, task-specific speech databases collected from realistic environments along with pronunciation in Bodo languages. The corpus for each language will consist of 100 speakers each contributing 200 tokens.
- **e.** Integration of speech interface, ASR engine and computer telephony interface
- **f.** Proposed a simple grid search based cluster model interpolation algorithm for model mean and/or mixture-weight adaptation which provides about 11% relative improvement in baseline performance of the recognizer module.

1.11 ORGANIZATION OF THE THESIS

In literature we found that there are five different components is spoken dialogue system and each component has several issues where researcher can put their contribution. As our research focus on word boundary detection, feature extraction approach in mobile network, system building and speaker adaptation hence we give the literature review chapter wise. The rest of the thesis is organized as follows:

CHAPTER 2 introduces the fundamental concepts of spoken dialogue system. We start with the history of spoken dialogue systems, its importance and its architecture. It also explains its classification, evaluation methodology, error handling in spoken dialogue systems, and gives some examples of available spoken dialogue systems. Different approaches to build up a spoken dialogue system and its different advantages and disadvantages are also described in this chapter.

CHAPTER 3 is devoted to the word boundary detection of Bodo language. Phonetically balanced, selected sentences are analyzed in time and frequency domain and the pitch contour of the respective sentences are thoroughly discussed. The pitch variation across a word in a sentence is measured and the differences of the last pitch value of a word and the first pitch value of the next word in a sentence is applied for word boundary detection. This chapter explains the experimental work done with speech feature like pitch, intensity, pause, and silence, voiced and unvoiced sound required for world boundary detection.

CHAPTER 4 proposes the Linear Prediction Coding and its algorithm, Mel Frequency Cepstral Coefficients algorithms and its implementations and simulation results for feature Extraction and also discusses the speech recognition approaches like Dynamic time warping and Hidden Markov Model Dynamic time warping and its implementations and simulation results are presented.

CHAPTER 5 explains the design and development of spoken dialogue system along with speech recognizer for accessing agricultural commodity information. This Chapter covers how the various algorithms and data structures presented in this thesis are technically integrated in the system architecture. We explain how the PHP AGI Script with HTK toolkit is structured in ASTERISK server. Finally, the chapter presents the integrated dialogue system employed to carry out the experiments in this thesis

CHAPTER 6 presents the analysis of the performance of the developed system, in terms different parameters like gender based analysis, WER, handset variability, VAD analysis, Burgin analysis etc. Based on small training set of Wizard-of-Oz interactions, the evaluation contrasted the empirical performance using baseline model, adaptation based model. The empirical results showed that the use of adaptation based models yields significant improvements in both objective and subjective metrics of interaction quality compared to the two baselines.

CHAPTER 7 presents the concluding remarks based on the present study. Finally, the possible benefits which can be acquired by further extending the present study have been discussed.