7

CONCLUSION AND FUTURE SCOPE

CONTENTS

7.1 CONCLUSION

The interaction between human and computer try to minimize the gap for literate/illiterate and visually challenged users to access information. There are different kinds of ways like speech, text, gestures, symbols etc., or a combination of these for interaction between human and computer. An interaction or conversions consist of statements, expressions, questions and answers. A spoken dialogue system is computer agent that interacts with people by understanding spoken language. The goal of a spoken dialogue system is to provide information by conversing with a human-being in a natural fashion. The main purpose of a spoken dialogue system is to provide an interface between a user and a computer-based application such as a database or expert system.

Though computer is the most popular and effective means to access information and helps to make our work easier. But still it requires a skill to operate the computer. It's a great challenge for physically handicapped or blind people to operate computers. Thus, speech synthesis and speech recognition systems play a vital role in such scenarios. This thesis discusses some of the steps required for the automatic segmentation of speech. There are various characteristics of speech such as: voiced, unvoiced and silence. Such basic characteristics of the speech can be evaluated by the computation of zero crossing rate, short term energy fundamental frequency, energy etc. All these characteristics are analyzed in this thesis. The acoustic signal is segmented into some basic units. The syllables are one of the most important units of segmentation. The function STE contains useful information about the peaks and valleys thus, helps to define the segment boundaries. The peak having maximum value called the nucleus is represented as vowel is represented as consonants. In our work STE is found to be the best method for speech segmentation. The higher value of short term energy refers to the voiced segments. The valleys at both the ends represent the boundary of the syllable. ZCR can only categorize the signal into voiced or unvoiced whereas the Short term energy can compute the energy content of the signal and also helps to mark the syllable boundaries by detecting the peaks and valleys in the speech signal.

In this thesis the basics of speech recognition system and different approaches available for feature extraction and pattern matching has been discussed. Using these various techniques rate of speech recognition can be improved and better quality

speech recognition can be developed. By combing the different features with different pattern matching techniques we found the MFCC with hmm gives better recognition accuracy in online isolated word recognition over telephone network. In future there will be focus on development of large vocabulary speech recognition system and speaker independent continuous speech recognition system. For developing such systems in future Artificial Neural Network (ANN) and Hidden Markov Model (HMM) will be used at high level as in recent these techniques have become popular techniques in speech recognition process.

The development of a spoken dialogue system for accessing the information of agricultural commodities in Bodo language is described in this thesis. The main goal was to integrate a telephony server with a spoken dialogue system for accessing agricultural commodity information. Asterisk PBX was used to act as the telephony server for this system. For communicating with user caller transfer their caller via Bluetooth. That enabled us to connect to the system using different devices such as land-line phones, cellular phones, and IP phones. We faced multiple problems related to the system that were solved using some custom dial plans. The secondary goal of this thesis was to integrate this spoken dialogue system with the Agri information service. A dialogue manager was Implemented using php-agi script for this purpose. This dialogue manager had two tasks; the first one was to act as a link between the system and service. Using Bluetooth as a connecting medium it provides services to the user. The second purpose of this dialogue manger is to handing error if caller make some mistake.

This developed model is cost effective and a good solution for research purpose. Although the system is developed using open source tools available following the standard procedures, there are some contributions made in terms of Bodo language, ASR integration with speech recognizer, voice-unvoiced-silence classification, boundary detection, analysis of different speech features for recognition of Bodo word in real time environment with mobile network, analysis of collected speech corpus etc. A metric for quantifying the user's comfort level is also proposed and the developed SQ system is found to result in an average user's comfort level of 69%. We have discussed the baseline agricultural information system and also an improved version using multiple decoders and contextual information. The issues in on-line adaptation with unseen speaker and small data in the context of the developed

SQ system are also discussed in this work. Two simple acoustic model interpolation based adaptation approaches have been explored to improve the system performance.

7.2 FUTURE SCOPE

The biggest challenge for future is to achieve as human-human like conversation as possible. The future system should be capable of handle “barge-in” and switch the context between system and user. Ideally system should allow user to pass information as flexibly as possible, without setting any restrictions on what, when and how user do it. One interesting question is how to make user expectations to match the machine capabilities. System should be able to recognize when user is trying to do something out of scope. Otherwise user might think that system has misunderstood him. And when the system really interprets something incorrectly, how to recognize it and inform user before anything harmful happens. And how could system learn from the error and not to do it again?

In speech recognition and language understanding there are many things to improve. One can also always improve the output of the system. The way the system responds to the user affects strongly on the overall impression of the system and its intelligence. Spoken dialogue systems offer some obvious benefits. It is in everybody’s interest that we can use computers more effectively and flexibly; in a more human like way. It looks like that there will be lots of efforts put on this research area and substantial progress will be made in the coming years.

In this work the base line system for spoken dialog system for agricultural commodities in Bodo language developed and tested. The trainer and the decoder configuration files have several parameters; these parameters can be tested to improve the efficiency of the spoken dialog system. Hence following technique can be used to improve the recognition accuracy.

a. **Adaptation/Normalization**: The system that will be built will be speaker-independent system. However, we could investigate the use the vocal-tract length normalization (VTLN) to improve the accuracy. Of course, the emphasis should be on the real-time application of this method. Similarly, we can use some of the speaker-adaptation techniques (like MLLR) to improve the recognition.

b. **Environment Adaptation**: We could use multi-condition training i.e., speech collected under different noisy conditions to improve the recognition accuracy of the

system in noisy conditions. We would also study the use of adaptation methods to improve noise robustness or techniques like parallel-model combination.

c. Discriminative Training: Since recognizing commodity names is a challenging problem due to the acoustic similarities between the different commodities, we plan to investigate the use of discriminative training in improving the accuracy of this set.