An Intelligent Speech Recognition System for Education System

Vishal Bhargava, Nikhil Maheshwari
Department of Information Technology, Delhi Technological University (Formerly DCE), Delhi
vishalbharg@gmail.com , nik.dtu@gmail.com

ABSTRACT:

E-Learning is the fastest and inexpensive source of information today. Thousands of the people today are using electronic gadgets for accessing E-Books, Notes, News, Stocks, Entertainment and Education. E-Learning, Now a day is getting popularity because of its unlimited benefits. A person, sitting anywhere around the globe can access E-Resources from his home provided he doesn't have disability of any kind. Getting proper information is a difficult task and it is also a big challenge to blinds and other disabled persons [4]. So requirement of developing a Speech User Interface was felt.

Speech Recognition is a promising way for users to control computer applications, especially when the users are unable to use traditional input devices like keyboard and mouse.

The objective of this paper is to present an Automatic Speech Recognition model, which is speaker independent, speech dependent and based on isolated words with small vocabulary by recognition of phonemes, groups of phonemes and words. We develop an application which is working as English language basic tutorial.

Keywords: E-learning, AIML, Speech Recognition.

I. INTRODUCTION:

The proposed system presents an Automatic Speech Recognition model, which is speaker independent, speech dependent and based on isolated words with small vocabulary by recognition of phonemes, groups of phonemes and words using below specified techniques.

The brief of these techniques is necessary to understand the system better.

First, AIML (Artificial Intelligence Markup Language) [2] is an XML-based language which is easy to learn, and allow customizing previously available Alicebot or creating new one.

The most important units of AIML are:

- <aiml>: AIML document
- <category>: the tag that marks a "unit of knowledge" in knowledge base
- <pattern>: contain a simple pattern that matches what a user may say or type
- <template>: contains the response to a user input

There are approximately 20 additional tags often found in AIML files, and it's possible to create your own "custom predicates".

Second, The Hidden Markov Model Toolkit (HTK) [3] is a portable toolkit for manipulating and building hidden Markov models. Primary use of HTK is for speech recognition research although it is used for numerous other applications such as research into speech synthesis, recognition of characters and sequencing of DNA structure. Hundreds of sites worldwide are now using HTK.

HTK is made up of a set of library modules and tools available in C source form. This toolkit provides complete facilities for speech analysis, training of HMM, testing and analysis of results. The software support is available for HMMs using continuous density mixture Gaussians as well as for discrete distributions and is used to build complex HMM systems.

Third, Feature Extraction is a crucial step in Pattern Recognition.

Fourth, Proposed system do auto correction of spelling and put correct word on proper place.

This organization of this paper is as follows: Section II presents related work and research objectives. Description of the system and implementation is discussed in section III. Results are described in Section IV and finally conclusions and future work are drawn in section V.

II. RELATED WORK, RESEARCH MOTIVATION AND OBJECTIVES

In the present Windows Speech Recognition [8] available in Windows Vista and in Windows 7, empowers user to interact with their computer by voice. This API allows user to control their system by voice, or it can be voice recognition system by IBM used in Honda cars which delivers in-car speech-recognition navigation system but problem with these system is that they are pronunciation dependent and to make these system according to user’s pronunciation its take too long time to be trained. So these systems cannot work as multiuser system and behave only like a single user system. “E-learning system for Japanese Transitive and Intransitive Verbs” [9], “Mandarin e-learning system” [11] are few great initiative taken into the direction of learning foreign languages using speech recognition system.

III. SYSTEM DESCRIPTION

A. Overview and interface design

In the design and development of proposed system we have taken care of mainly two things. First, Presence of a friendly and rich GUI of the designed system. Fig. 1 shows the block diagram of our system.

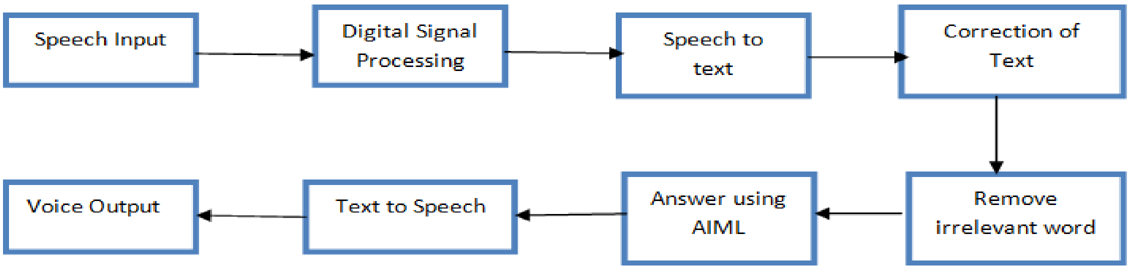

Fig. 1: Block Diagram of Proposed System

Second, Answers of different queries fired are stored in system itself in plain English and in proper length, so user gets their answer in easy and effective way. After ask the question its interface also display the asked question to the user, and given answer is also displayed on the screen.

We develop our system as e-learning system for the children, like they can ask the question – "What is noun?" so system answer them according to answer store in repository like "Noun is the name of person, place or thing". After give the answer our system also ask about the repeat of the answer if user say yes than it repeat, if user say no or default it doesn't repeat. In the continuation system also ask about do you want any example again depend on user's response it answer. Apart learning system also give general answer to the user like ask about direction, about institute.

B. EXPERIMENTAL SETUP & IMPLEMENTATION

The whole system is developed in C# and experiments are performed on the system.

I. Recording speech:

The First step in the system development is Recording of Speech. This is done by collecting the voice samples from different people with different accents. We have focused mainly college students and collected samples from the different colleges. These questions are in the form of queries[5] by the user as "Where is the administrative block?".

II. **Digital Signal Processing:**

Digital signal processing (**DSP**) [7] is used for the representation of signals by a sequence of numbers or symbols and the processing of these signals.

Digital Signal Processing is done on Speech Signal in following step:

1. Feature Extraction: Feature Extraction is done for extraction of phonemes from input speech signal. It includes Mel Frequency Cepstral Coefficient (MFCC) Algorithm. MFCC includes Pre-emphasis, Frame Blocking, Cepstral Co-efficient Extraction, parameter weighting.
2. Vector Quantization: The role of Vector Quantization is to match the speech phonemes with reference phonemes like dictionary phonemes.
3. Hidden Markov Model: The role of HMM [6] is to find out most probable path for input word signal using Viterbi Algorithm. The role of viterbi algorithm is to find out most probable word.

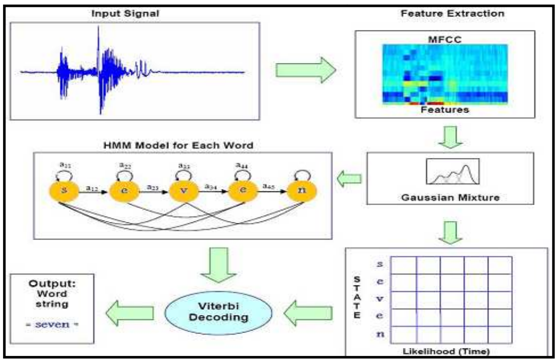

Fig. 2: Digital Signal Processing

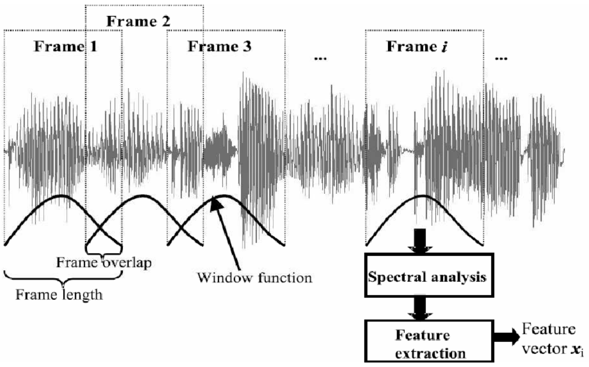

Fig. 3: Framing of Speech Signal

III. Speech To Text:

For speech to text conversion, .NET predefined library SpeechLib of SDK 5.1[12] is Used.

IV. Correction of Text:

The proposed system is an intelligent system which has a predefined vocabulary. This vocabulary is used for correction of text [10]. The accent and pronunciation of different users is different so detection of words may also differ. System uses a language model & error-correcting algorithms to guess the word intended. It also includes a tapping predictive text system in the same interface. Eg. ‘Where’ can be detected as ‘were’, ‘wheeere’ or ‘weeere’ at the time of recognition. System does following to detect the correct word:

1. Check the word from vocabulary and correct the word. Similarly ‘wheeeere’ and ‘weeere’ are no logical words so cancelled.
2. On the basis of predefined question structure, it is fixed that ‘were’ cannot be the first word of any query so it is cancelled.

V. Remove Irrelevant Words:

In this phase, Removal of irrelevant words is done. In the process of correction of text many irrelevant data or words are added in the query e.g. ‘the’ and removal of such words will not affect the meaning of query.

So now the question will be like “Where administrative Block?”

VI. Artificial Intelligence Markup Language (AIML):

To answer the query, AIML[2] is used.

e.g.

```
<category>
         <pattern>WHERE ADMINISTRATIVE BLOCK</pattern>
         <template>
         <think><set name="topic">Me</set></think>

        THIS IS RIGHT SIDE FROM FIRST CROSSING AFTER MAIN GATE.

        </template>
</category>
```

The category section contains different knowledge units within it. These units are basically different queries. These queries are written in pattern section and answer to that query is written in template written just below it. Thousand of such queries and their answers are present in consecutive patterns and templates.

VI. Text To Speech:

The Text written in template is then converted back to speech by SpeechLib Library of .Net Speech SDK 5.1.

VII. Speech Output:

The Speech generated is produced as output at the UI.

IV. RESULTS

The performance of the system was measured by computing the recognition accuracy at the word level.

The word accuracy can be shown as

$$\text{Word Accuracy} = \frac{100^{*}\ \text{No. of Words Correctly Recognized}}{\text{Total number. Of words in the test suite}}$$

We used HTK toolkit to compute the accuracy by comparing the hypothesized words and the actual word.

The Table 1 shows the Performance Speech Recognition System

Type of data	**No. of words**
Training	2000
Test	200
Recognized Words	156

Table 1. Performance Speech Recognition System

In our system Accuracy (%) words = 78%

The performance of the system can be improved by implementing Feature extraction techniques for noisy environment. General observations for improvement of performance are as following:

- Train the speech recognition system in the implementation environment.
- Keep vocabulary Size as small as possible
- Keep short of each speech input (word length).
- Use speech inputs that are distinctly different from each other.
- Provide immediate echo for each speech input.
- Keep the user interface simple.
- Error correction should be intuitive.
- Add functionality to quickly and easily turn off and on the speech recognizer.
- Use a directional, noise-canceling microphone
- Use headphones or an earphone for auditory feedback.

V. CONCLUSION AND FUTURE WORK

One challenging application which may revolutionize the way we use our system is to make it pronunciation independent in the easiest manner. Proposed System works with the intension of fulfilling this requirement to some extent. In this paper we have proposed a system that is useful not only as a tool for pronunciation independent system, but also as an effective means for retrieving knowledge and finding answer of interest, specially for disables.

Currently, proposed system is providing service in our college, but it can also be implemented on other public place like at the time of games, events etc. In future to give answer to the user we can develop our own framework which can provide answer in Indian tone or according to user specific tone, in place of current Microsoft APIs. The major problem in this area is environmental noise that reduces the performance of the system. New methods and algorithms are coming up for feature extraction, noise filtering and distorted signal enhancement. Some more techniques for better performance are also required like echo cancellation, pitch normalization, energy, gain improvement etc. Ambiguity in question, like Ambiguity in verb and adjective is also a big challenge now and it needs improvements.

REFERENCES

1. J. Ramírez, et al, Voice Activity Detection. Fundamentals and Speech Recognition System Robustness, ISBN 987-3-90213-08-0, pp.460, I-Tech, Vienna, Austria, June 2007.
2. http://www.alicebot.org/aiml.html
3. http://htk.eng.cam.ac.uk/
4. Li Bian , ‘E-Learning with Computer Technology in Handicapped Higher Education, Coll. of Special Educ’., Beijing Union Univ., EBISS '09 Beijing , May 2009
5. Thorsten Brants , “Natural Language Processing in Information Retrieval”,Google Inc, 2003.
6. Mikael Nilsson, Marcus Ejnarsson, Speech Recognition using Hidden Markov Model, MEE-01-27
7. Rabiner L, Juang B H, Fundamentals of Speech Recognition (New Jersey, USA, Prentice Hall, 1993).
8. http://www.microsoft.com/enable/products/windowsvista/speech.aspx
9. Okumoto, H. (2004). Japanese Transitive and Intransitive Verbs: e-Learning System with Speech Recognition and Video. In L. Cantoni & C. McLoughlin (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2004 (pp. 1902-1907). Chesapeake, VA: AACE.
10. Joseph J. Pollock, Antonio Zamura “automatic spelling correction in scientific and scholarly text” ACM ISBN 0001-0782/84/0400-0358, Apr 84.
11. Yue Ming, Zongshan Bai “A Mandarin e-learning system based on speech recognition and evaluation” Wiley Periodicals, Inc 2009.
12. http://www.microsoft.com/downloads/details.aspx?FamilyId=5E86EC97-40A7-453F-B0EE-6583171B4530&displaylang=en