An efficient online examination system using speech recognition

Aishwarya Rai[1], Arbab Khan[2], Abhishek Bajaj[3], Jasmine Bedi Khurana[4]

Dept. of Information Technology, Galgotias College of Engineering and Technology, Uttar Pradesh, India [1],[2],[3]
Assistant Professor-2, Dept. of Information Technology, Galgotias College of Engineering and Technology, Uttar Pradesh, India [4]

---***---

Abstract - *A speech recognition system converts the speech-sound into the corresponding text. The uttered speech is first understood by the machine and then the corresponding text is displayed [1]. This paper describes the development of an efficient speech recognition system for speaker independent, real time, isolated words of English Language. Speech is a useful and efficient communication medium with machines, especially in the environment where keyboard input is tedious or impossible. This paper is a study of the technology and modeling techniques have been used for Speech Processing and the application of ASR in Online Examination System. Online Examination System using Speech Recognition is very useful for Educational institutions to prepare an exam, save the time that will be needed to check the paper and prepare mark sheets. It will be particularly of help for the students with disabilities to appear for competitive examinations as well as university examinations. The integration speech recognition system will make the procedure of exam conducting, monitoring and completion more efficient, organized and less chaotic and extremely convenient.*

***Key Words*: Speaker independent, isolated, ASR , Online examinations , disabled students.**

1. INTRODUCTION

Speech recognition refers to the ability to listen (input in audio format) spoken words and identify various sounds present in it, and recognize them as words of some known language. Speech recognition in computer system domain may then be defined as the ability of computer systems to accept spoken words in audio format - such as the steps required to make computers perform speech recognition are: Voice recording, word boundary detection, feature extraction, and recognition with the help of knowledge models.[2] Word boundary detection is the process of identifying the start and the end of a spoken word in the given sound signal. This can be attributed to various accents people have, like the duration of the pause they give between words while speaking. Knowledge models refer to models such as acoustic model, language models, etc. which help the recognition system. To generate knowledge model one needs to train the system by providing voice samples to the corpus. There are many publicly available software tools for the research work in the field of speech recognition such as Sphinx from Carnegie Mellon University (SPHINX, 2011), hidden Markov model toolkit (HTK, 2011) and large vocabulary continuous speech recognition (LVCSR) engine Julius from Japan (Julius, 2011). This paper aims to develop and implement speech recognition with the help of Microsoft speech API (based on the HMM model) integrated in Microsoft Visual Studio using .net and C sharp coding.

2. SPEECH RECOGNITION

Speech recognition is the process of mapping an acoustic waveform into a text (or the set of words) which should be equivalent to the information being conveyed by the spoken words.

2.1 Components of Speech Recognition System

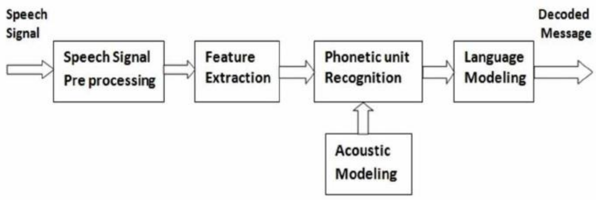

Fig- 1: Modules of speech recognition system

2.1.1 Speech Signal acquisition: At this stage, analog speech signal is acquired through a high quality, noiseless, unidirectional microphone in .wav format and converted to digital speech signal.

2.1.2 Feature Extraction: Feature extraction is a very important part of SR system development during which a frugal sequence of feature vectors is computed so as to provide a compact representation of the given input signal. Speech analysis of the speech signal acts as first stage of Feature extraction process where raw features describing the envelope of power spectrum are generated.

2.1.3 Acoustic Modeling: Acoustic models are developed to link the observed features of the speech signals with the expected phonetics of the hypothesis word/sentence.

2.1.4 Language & Lexical Modeling: Word ambiguity is an aspect which has to be handled cautiously and acoustic model alone cannot handle it. Lexical model supplies the pronunciation of the words in the specified language and contains the mapping between words and phones [3],[4] Generally, a canonical pronunciation available in traditional or standard dictionaries is used. The main idea of performing adaptation is to minimize the systems performance dependence on speaker's voice, microphones, transmission channel and acoustic environment so that the generalization capability of the system can be enhanced.

2.1.5 Recognition: Recognition is a mechanism where an unknown test sample is contrasted with each sound class reference pattern and, thereby, a measure of similarity or closeness is calculated.

2.2 Fundamentals to Speech Recognition

Speech recognition is basically the science of talking with the computer, and having it correctly recognized [9]. To elaborate it we have to understand the following terms [8], [11].

2.2.1 Utterances

When user says some things, then this is an utterance [11] in other words speaking a word or a combination of words that means something to the computer is called an utterance. Utterances are then sent to speech engine to be processed.

2.2.2 Pronunciation

A speech recognition engine uses a process word is it's pronunciation, that represents what the speech engine thinks a word should sounds like. Words can have the multiple pronunciations associated with them. Language model predicts the probability of a word occurring in the context. In some cases, it may be possible that there are words which are phonetically similar but having different meaning called homophones. Handling such homophones is a vital issue for any ASR, as they generally increase acoustic confusability. Different languages have different number of homophone words. For example, French language admits a large number of homophones; hence in this particular case automatic transcription is a very challenging task. Homophones are also prevalent is English language (like I or eye, week or weak, principal or principle). Speech system recognizes these words based on the context in which they occur. Language model provides this context to the speech recognition system. There are mainly four approaches to language modeling viz. grammar-based approach, stochastic approach, uniform modeling and finite state approach.

Hidden Markov Processes are the statistical models in which one tries to characterize the statistical properties of the signal with the underlying assumption that a signal can be characterized as a random parametric signal of which the parameters can be estimated in a precise and well-defined manner. In order to implement an isolated word recognition system using HMM, the following steps must be taken:

(1) For each uttered word, a Markov model must be built using parameters that optimize the observations of the word.

(2) Maximum likelihood model is calculated for the uttered word. [5], [9], [10], [11].

2.2.3 Grammar

Grammar uses particular set of rules in order to define the words and phrases that are going to be recognized by speech engine, more concisely grammar define the domain with which the speech engine works. Grammar can be simple as list of words or flexible enough to support the various degrees of variations.

```
Choices commandChoices = new Choices("A", "B", "C", "D", "E", "F", "G", "H", "I", "J", "K", "L", "M", "N", "O", "P", "Q", "S", "T", "U", "V"
GrammarBuilder grammarBuilder = new GrammarBuilder();
grammarBuilder.Append(commandChoices);
Grammar g = new Grammar(grammarBuilder);
g.Name = "Available programs";
```

Fig-2: Grammar builder

2.2.4 Accuracy

The performance of the speech recognition system is measurable [8]; the ability of recognizer can be measured by calculating its accuracy. It is useful to identify an utterance.

Table 1: Recognition accuracy

Word	Recognition%
Male	80%
Female	85%
Computer	85%
Robot	78%

Word Error Rate
By contrasting the number of words in transcription of the test data which has been recognized incorrectly by your system compared to the transcription you will make manually.

Sentence Error Rate
By contrasting the number of sentences in transcription of the test data which has been recognized incorrectly by your system compared to the transcription you will make manually.

2.2.5 Vocabularies

Vocabularies are the list of words that can be recognized by the speech recognition engine. Generally the smaller vocabularies are easier to identify by a speech recognition engine, while a large list of words are an arduous task to be identified by the engine.

2.2.6 Training

Training can be of help to the users who have difficulty in speaking or pronouncing certain words. Speech recognition systems with training should be adaptable.

3. PROPOSED MODEL

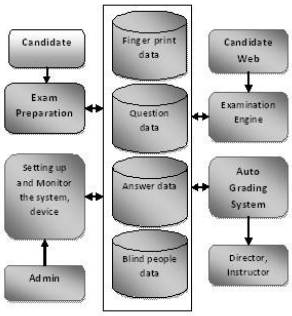

Fig- 3: Architecture of the proposed model

Database Tier

Here data are stored as records in tables on the server side database.

Exam system data

This contains all question data; all answer data, these data are explained as follows:

Exam Preparation:

It is used to manage and handle the questions asked in the exam. It also contains the logic behind instructor-course relation, instructor-term relation and the report behind these relations. This logic used to handle all the information stored in the database about admin relation. A detail of that logic is described as follows:

Add questions: The administrator could first insert the questions in the database record.

Create exam: The admin could create exam by selecting the questions that added before.

Update exam: The admin could update the exam that made before.

Schedule exam: The admin can schedule the exam for the blind student.

Setting up and monitoring the system: It is used to set and handle the candidate information which is detailed as follows: The administrator login, he can insert or update candidate information. There are some scenarios for that event. The administrator login and he can insert or update instructor information he can check for the instructor first if the instructor does not exist the administrator can create new instructor record by adding instructor data.

Auto grading: With that, our system can automatically grade candidates' answers, which are collected by the examination system. The system compare the student answers with the correct answers which entered by the admin.

4. TOOLS USED

SAPI

The Speech Application Programming Interface or SAPI is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. Broadly the Speech API can be viewed as an interface which sits between applications and speech engines (recognition and synthesis).

SAPI 5.1

This version was launched in late 2001 as part of the Speech SDK version 5.1. Automation-compliant interfaces were added to the API to allow use from Visual Basic, scripting languages such as JavaScript, and managed code. This version of the API and TTS engines was shipped in Windows XP.[10] Windows XP Tablet PC Edition and Office 2003 also include this version.

```
using System;
using System.Collections.Generic;
using System.ComponentModel;
using System.Data;
using System.Drawing;
using System.Linq;
using System.Text;
using System.Windows.Forms;
using SpeechLib;
using System.Speech.Recognition;
using System.Data.SqlClient;


namespace AptitudeTest
{
    public partial class LoginPage : Form
    {
        SqlConnection cnn = new SqlConnection(Program.con());
        SpFileStream spFileStream = new SpFileStream();     //declaring and Initializing fileStream obj
        SpeechStreamFileMode spFileMode = SpeechStreamFileMode.SSFMCreateForWrite;  //declaring fileStreamMode as to Create or Write
        SpObjectTokenClass my_SpObjToken = new SpObjectTokenClass();    //declaring and initializing SpObject Token Class
        SpeechVoiceSpeakFlags my_Spflag = SpeechVoiceSpeakFlags.SVSFlagsAsync; // declaring and initializing Speech Voice Flags
        SpVoice my_Voice = new SpVoice();                   //declaring and initializing SpVoice Class
        int sp_Rate = -250, sp_Volume = 70;
```

Fig-4: Speech libraries integrated in windows platform

Microsoft SQL Server 2008

This platform has been used to for the back end development of this system where the entire database of the proposed system will be created and managed systematically.

5. CONCLUSION

This project would be a very useful one for every blind or specially-abled student, so that we can examine their capabilities easily and more effectively through an online exam like every other normal student. And also we will try to do as much refinement in future as per the collection of feedback. One of the major goals for the future will be to include Hindi and other native languages so that the system can be enlarged to support the illiterate as well.

REFERENCES

[1] "Speech recognition- The next revolution" 5th edition.

[2] Forgie, J. W. and Forgie, C. D. (1959) 'Results obtained from a vowel recognition computer program, *J. Acoust. Soc. Am.*, 31(11), pp.1480-1489.

[3] Hemdal, J.F. and Hughes, G.W., A feature based computer recognition program for the modeling of vowel perception, in Models for the Perception of Speech and Visual Form, Wathen-Dunn, W. Ed. MIT Press, Cambridge, MA.

[4] Watcher, M. D., Matton, M., Demuynck, K., Wambacq, P., Cools, R., "Template Based Continuous Speech Recognition", IEEE Transaction on Audio, Speech, & Language Processing,2007.

[5] Watson, B. and Chung T. (1992) 'Second order Hidden Markov Models for speech recognition'. *Paper Presented at Fourth Australian International Conference on Speech Science and Technology*, pp.146-151.

[6] Polur, P. D., Zhou, R., Yang, J., Adnani, F. and Hobson, R. S. (2001) 'Isolated speech recognition using artificial neural networks'. *Paper Published in Proc. of the 23rd annual EMBS international conference,* IEEE, Turkey, pp.1731-1734.

[7] Runstein, F. and Violaro, F. (1995) 'An isolated-word speech recognition system using neural networks'. Paper published in the Proceedings of the 38th Midwest Symposium on Circuits and Systems, IEEE, pp. 550-553.

[8] Dev, A. (2009) 'Effect of retroflex sounds on the recognition of voiced and unvoiced stops', *Journal of AI and Soc.*, Springer, Vol. 23, pp. 603-612.

[9] Niles, L. and Silverman, H. (1990) 'Combining Hidden Markov Models and Neural Network classifiers'. *Paper Presented in ICASSP*, pp.417–420.

[10] Rigoll, G. (1994) 'Maximum Mutual Information Neural Networks for Hybrid connectionist-HMM speech Recognition Systems', *IEEE Transactions on Audio, Speech and Language processing*, Vol. 2, No.1, Part II.

[11] Jain, R. And Saxena, S. K., "Advanced Feature Extraction & Its Implementation In Speech Recognition System", IJSTM, Vol. 2 Issue 3, July 2011.

BIOGRAPHIES

Ms. Aishwarya Rai studying I.T in Galgotias College Of Engineering and Technology , Greater Noida , India.