ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

SPEECH RECOGNITION SYSTEM FOR ENGLISH LANGUAGE

Ms. Vrinda[1] ,Mr. Chander Shekhar[2]

Astt.Professor, Deptt of Computer Science & Engineering, M. V N University, Palwal, India[1]

Manager -IT, Technical Textile Business, SRF Limited, Gurgaon, India[2]

***Abstract*- This paper presents an overview of speech recognition technology. This report presents an overview of speech recognition technology, software, development and applications. It begins with a description of how such systems work, and the level of accuracy that can be expected. Present work is aimed at developing suitable speech recognition for hindi language, for the people who are physically hampered and cannot able to operate the computer through keyboard and mouse. In this paper we are using a HMM (hidden Markov model) to recognize speech samples to give excellent results for isolated words. It consists of isolated words that are separated by silences. The advantage of discrete speech is that word boundaries can be set exactly while with continuous speech; words will be spoken without silences.**

Keywords- Speech Recognition, HMM, Discrete Speech

I. INTRODUCTION

Speech is the primary means of communication between people. Speech recognition, generation of speech waveforms, has been under development for several decades [10]. Automatic speech Recognition is a process by which a computer takes a speech signal and Converts it into words [1]. It is the process by Which a computer recognizes what a person Said. Keyboard, although a popular medium, is not very convenient, as it requires a certain amount of skill for effective usage .A mouse on the other hand requires a good hand eye co-ordination. Physically challenged people find computer difficult to use. Partially blind people find reading from a monitor difficult. All these constraints have to be eliminated. Speech interface help us to tackle these problems.
The objective is to trap human voice in a digital computer and decode it into corresponding text. Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words.

When two people speak to one another, they both recognize the words and the meaning behind them. Computers, on the other hand, are only capable of the first thing: they can recognize individual words and phrases, but they don't really understand speech in the same way as humans do. Computer recognizes the command and software tells the computer what to do when that command is recognized.

II CLASSIFICATION OF SPEECH RECOGNITION SYSTEM

There is a large variety in the speech recognition technology and it is important to understand the differences. One can classify speech recognition systems according to the type of speech, the size of the vocabulary, the basic units and the speaker dependence[2]. The position of a speech recognition system in these dimensions determines which algorithm has to be used.

A. Type of speech

There are basically two types of speech:
1. Continuous speech
2. Discrete speech.

Discrete speech consists of isolated words that are separated by silences [3]. The advantage of discrete speech is that word boundaries can be set exactly while with continuous speech; words will be spoken without silences.

B. Size of the vocabulary

The size of the vocabulary is the second typical aspect of a speech recognition technology. The vocabulary is a set of words that have to be recognized. A small vocabulary is one, which contains less than about 30 words. A 500-word vocabulary is average size. A vocabulary with more than

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

25000 words generally will be seen as very big, although these definitions tend to depend on the application field [6**].**

C. Speaker dependence

- Speaker dependent system
- Speaker independent system
- Speaker adaptable system

Some speaker-dependent systems require only that the user record a subset of system vocabulary to make the entire vocabulary recognizable. A speaker-independent system does not require any recording prior to speaker-dependent system requires that the user record an example of the word, sentence, or phrase system use. A speaker independent system is developed to operate for any speaker of a particular type (e.g., American English). A speaker adaptive system is developed to adapt its operation to the characteristics of new speakers[5].

III DESIGN OF THE SYSTEM

The prepared system if visualized as a block diagram will have the following components: Sound Recording and word detection component, feature extraction component, speech recognition component, acoustic and language model.

A. Sound Recording and Word detection component

The component is responsible for taking input from microphone and identifying the presence of words. Word detection is done using energy and zero crossing rate of the signal. The output of this component can be a wave file or a direct feed for the feature extractor.

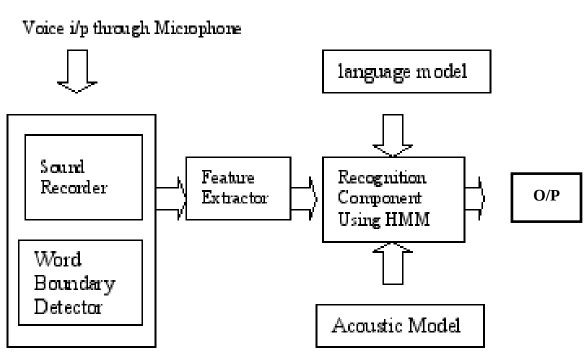

Figure 1. Block diagram of Training System

B. Feature Extraction component:

The component generated feature vectors for the sound signals given to it. It generates Mel Frequency Cepstrum Coefficients and Normalized energy as the features that should be used to uniquely identify the given sound signal [5].

C. Recognition component:

This is a Continuous, Multi-dimensional Hidden Markov Model based component. It is the most important component of the system and is responsible for finding the best match in the knowledge base, for the incoming feature vectors

D. Knowledge Model:

The components consist of Word based Acoustic. Acoustic Model has a representation of how a word sounds. Recognition system makes use of this model while recognizing the sound signal.

Once the training is done, the basic flow can be summarized as the sound input is taken from the sound recorder and is feed to the feature extraction module. The feature extraction module generates feature vectors out of it which are then forwarded to the recognition component. The recognition component with the help of the knowledge model and comes up with the result. During the training the above flow differs after generation of feature vector. Here the system takes the output of the feature extraction module and feeds it to the recognition system for modifying the knowledge base[11].

IV RESULTS

The system tested for various parameters and get the following result as follow. In the experiment Hidden Markov Model Is used for the recognition of isolated English words**.**

A. Training

To train the system we used 10 users. Each user trained the system for some time. Training of the system was done in very peaceful environment so that recognition can be increased

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 1, January 2013

B. Recognition

Recognition was tried on two kinds of sounds. Known user: The user whose voice used for training. Unknown: The user whose voice did not used for training. The result of experiment as shown in table 1

TABLE 1. RESULT OF RECOGNITION

Type of user	**No. of sound**	**Correct recognition**	**Incorrect recognition**
Known user	10	8	2
Unkown user	10	5	5

TABLE 2: WORD RECOGNITION PROBABILITY

Words	**Number of testing**	**Recognition Probability**
Station	10	100%
Student	10	100%
Try	10	100%
Lemon	10	100%
Mobile	10	90%
Biology	10	80%

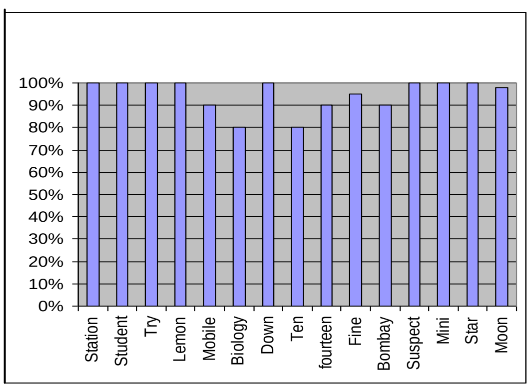

Figure 2 : Word Recognition Probability for 10 users

C. Implementation

In this project a word acoustic model is used .The system has a model for each word that system can recognize . In this model the words are modeled as whole. The language model contains list of words. While recognizing the system need to know where to locate the model for each word and what word the model correspond to .This information is stored in a flat file called models in a directory .These models in a directory are known as HMMs.If user need to add support to recognize a new word, user will have to train the system for the word. In the recognition process, the sound is matched against each of the model to find best match. This best match is assumed to be the spoken word. Building a model for a word requires us to collect the sound files of the word from various users. These sound files are then used to train a HMM Model. Firstly, user train the system. Train command is used to train the system for a new word.

D. Experiments

The accuracy of a speech recognition engine is measured by its word error rate The isolated English word Recognition system consist of five hundred Isolated words. Testing of randomly chosen fifty words is made by different speakers and the results are as follows

TABLE 3. RESULTS SHOWING NO. OF SPEAKERS VS ACCURACY

Down	10	100%
Ten	10	80%
Fourteen	10	90%
Fine	10	95%
Bombay	10	90%
Suspect	10	100%
Mini	10	100%
Star	10	100%
Moon	10	98%

No. of speakers	**Accuracy**
First	60%
Second	70%
Third	75%
Fourth	50%
Fifth	65%
Sixth	70%
Seventh	80%
Eighth	85%
Ninth	90%
Tenth	93%

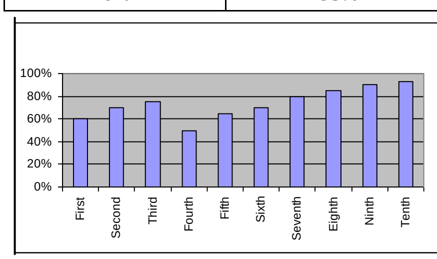

Figure 3. Accuracy Vs Number of speakers

The speech recognition system is a speaker independent system and is tested by a number of speakers and every time result varies. So its accuracy varies from speaker to speaker. Figure 3 shows the accuracy calculated when the system is tested by ten different speakers.

E. Accuracy

Accuracy of the experiments depends upon the training time .If the Training time of the system is increased ,accuracy automatically increased. time is directly proportional to accuracy .If training time increases then accuracy will increase automatically.

V AREA OF APPLICATIONS

This software is ideal for people who have a lot of documentation work to do. According to Samuel Dania of Philips Speech processing (which manufacture Philips Free Speech voice recognition software), people choose speech recognition software because it helps finish work faster[8]. Speech recognition software has improved to such an extent that one can use it to create reports, fill forms and use them as transcribers[9]. For doctors, lawyers and other professionals who need to generate reports everyday, this is quite a boon.

VI CONCLUSION

In this paper we studied how to trap human voice in a digital computer and decode it into corresponding text. Through this paper, we present a scheme to convert speech to text. The key factor in designing such system is the target audience. For example, physically handicapped people should be able to wear a headset and have their hands and eyes free in order to operate the system. In this Hidden Markov Model and various technique are used .A word based acoustic model is used. This model can be used only for limited vocabulary. As the size of the vocabulary increases performance of the system decreases. . The system cannot properly distinguish between similar words. Like to and two because they have similar sound phonemes. At last we conclude that this project can be used at very large scale with very little modifications.

During the experiment work medium size vocabulary system was implemented. The system can be extended to continuous word recognition with large vocabulary based on a phone acoustic model, using the HMM Technique or using other growing techniques like Artificial Neural Network.

ACKNOWLEDGMENT

We would like to thanks to all my friends who support me in writing this paper.

REFERENCES

[1] Anne Johnstone Department of Artificial Intelligence Edinburgh University Hope Park Square, Meadow Lane Edinburgh EHB 9LL, (GB) Gerry Altmann "AUTOMATED SPEECH RECOGNITION: A FRAMEWORK FOR RESEARCH".

[2] Reddy, D.R. & Ermann, L.D. 1975. "Tutorial on System Organisation for Speech Understanding." In D.R. Reddy (ed) Speech Recognition, Academic Press.

[3] Rumelhart, D.E. & McClelland, J.L. 1982. "An Interactive Activation Model of Context Effects in Letter Perception: Part II. The Contextual Enhancement Effect. Some Tests and Extensions of the Model. In Psychological Review"

[4] Tetsuya Matsumoto, Kazuhito Hagio, and Masayuki Takeda Department of Informatics, Kyushu University, Fukuoka 819-0395, Japan {tetsuya.matsumoto, kazuhito.hagio, takeda "More Speed and More Compression: Accelerating Pattern Matching by Text Compression"

[5] C. H. Lee; F. K. Soong; K. Paliwal "An Overview of Speaker Recognition Technology", Automatic Speech and Speaker Recognition: Advanced Topics. Kluwer Academic Publishers 1996, Norwell, MA.

[6] "Pattern matching for large vocabulary speech recognition systems" available at www.freepatentsonline.com/6879954.html

[7] "Isolated-word automatic speech recognition (iwasr) design" available at www.dspace.fsktm.um.edu.my/xmlui/bitstream/ handle/1812/111/Chapter%205.pdf?sequence=7

[8] R. Rodman, "Computer Speech Technology". Artech House, Inc. 1999, Norwood, MA 02062.

[9] M. J. Castro; J. C. Perez, "Comparison of Geometric, Connectionist and Structural Techniques on a Difficult Isolated Word Recognition Task.", Proceedings of European Conference on Speech Comm. and Tech., ESCA, Vol. 3, pp 1599-1602, Berlin, Germany, 1993

[10] Rabiner Lawrence ,Juang Biing-Hwang, "Fundamental of speech recognition ",AT & T,1993.

[11] "Speech recognition for Hindi language", C-DAC India, available at http://www.cdacmumbai.in/design/corporate_site/override/pdfdoc/speech_recognition_for_hindi.pdf