ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

REAL TIME INFORMATION SYSTEM BASED ON SPEECH PROCESSING: RADIO STATION AS VIRTUAL EYE FOR THE BLIND

Ranu Dixit[1] Navdeep Kaur[2]

M.Tech Students, Information Technology, Chandigarh Engineering College, Landran, Mohali,Punjab, India[1]

Faculty of Information Technology, Chandigarh Engineering College, Landran,Mohali, Punjab, India[2]

ABSTRACT: Till date blind people struggle a lot to live their miserable life. Their problems have made them to lose their hope to live in this competing society. They seek help from others to guide them whole day. This paper aims to make the blind person fully independent. The main goal of this paper is help blind and visually impaired person to find information for their everyday lives when it needed through radio station. It involves the synchronization of different sound signals and reading of each signal bit wise. These signals store in database, apply the HMM algorithm and distribute each sound signal in one information database. Speech recognition system is applies and each sound signal must create sound corresponding to it in database. .Net simulation tool is use to implement our approach. We present experiments result show how effectively it helps blind and visually impair persons.

Keywords: Automatic Speech Recognition, Hidden Markov Model (HMM), .Net simulation tool, classifiers, feature extraction, performance evaluation, Database.

I. INTRODUCTION

A. Blind person and information technology: Sight is consider to be the most essential of all the senses, people lacking this sense are look upon with pity by others. Visually impair persons face numerous difficulties to perform their day to day task. They are totally or partially dependent on someone for help [1]. In order to overcome these drawbacks, we design a system to collect recent information through automatic speech recognition by radio at any time and providing the solution, which would be of immense advantage in comparison to the devices that are using by blind people today. This system is capable of fulfilling all the requisites of visually impair and help them to become independent like normal humans.

II. OBJECTIVE

A. Radio as a tool for information: Traditionally radio programs were broadcast on different frequencies via FM and AM, and the radio had to be tuned into each frequency, as needed. This used up a comparatively large amount of spectrum for a relatively small number of stations, limiting listening choice. For the relay of the program, first it is record and then need to editing, after that at a suitable time table this is relay to the respective frequency. In this process audience may listen to this program if they are free at the same time, if they are busy they may not listen to the program, to avoid this problem real time communication known as live broadcasting is useful, We present experiments result show the help blind and visually impaired to find information for their everyday lives.

Blind user ask for information (to radio station) and the radio station replies for asked information e.g. hospital or Institute. The user selects the option by speech either for hospital by hospital or Institute, one of them. This option acts as input data for machine.

In radio communication information such as sound is transform into an electronic signal which is applies to a transmitter. The transmitter sends the information's through space on a radio wave (electromagnetic wave). A receiver intercepts some of the radio wave and extracts the information-bearing electronic signal, which is converted back to its original form by a transducer such as a speaker.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

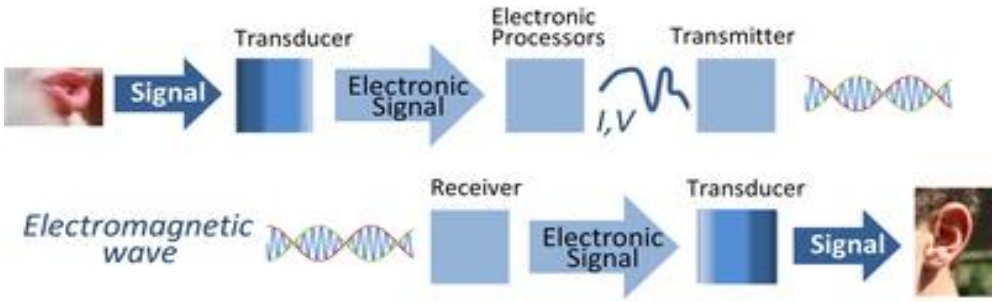

Figure 1: Radio communication steps

we must suppress interference to realize a noise-robust hands-free speech recognition the information of the observed signal in each microphone[2],[3],[4],[5],[6],[7],[8],[9],[10],[11],[12],[13],[14],[15],[16],[17],[18].Research in speech processing and communication for the most part, is motivate by people's desire to build mechanical models to emulate human verbal communication capabilities [19]. speech processing is one of the most exciting areas of the signal processing. Automatic speech recognition by machines has attracted a great deal of attention for sixty years [20].

III. DESIGN OF THE SYSTEM

Although the present software systems are often sophisticate and user-friendly, they are usually not very convenient for the visually impair people. The reason is the graphical interface and absence of the features fulfilling special needs of the visually impair. Speech synthesizer and screen reader software still represent basic functionalities that are used by the visually impair to obtain information by means of a computer [22].

The proposed system is visualized as a block diagram having the following components:

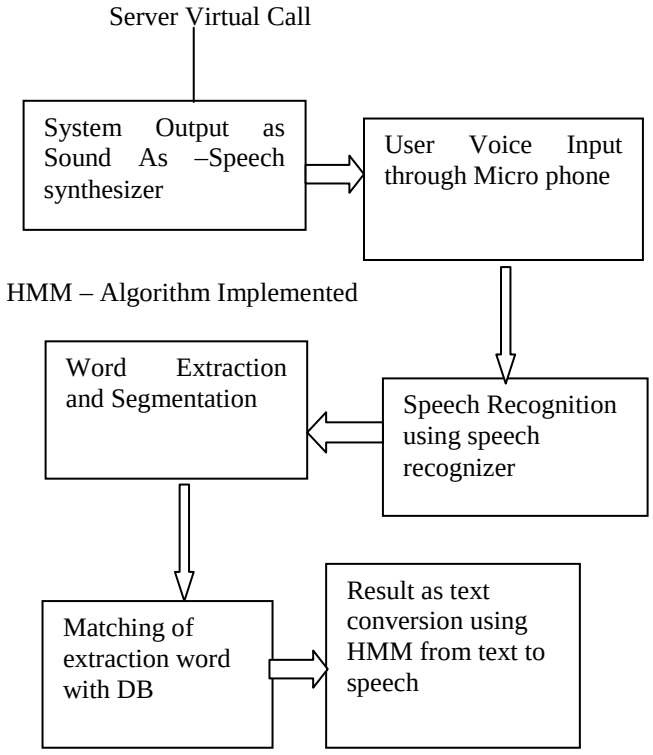

Figure 2: Block diagram of speech recognition

1. Sound Recording and word detection component
2. Feature extraction component Speech recognition component
3. Acoustic and language model.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

A. Sound recording and word detection: The component responsibility is to accept input from a microphone and forward it to the feature extraction module. Before converting the signal into suitable or desire form it also does the important task of identifying the segments of the sound containing words. It also has a provision of saving the sound into WAV files which are needed by the training component. The recorder takes input from the microphone and saves it or forwards it depending on which function is invoked Recorder supports changing of sampling rate, channels and size of the sample.

Word detector

In speech recognition it is important to detect when a word is spoken. The system detects the region of silence. Anything other than silence is consider as a spoken word by the system. The system uses energy pattern present in the sound signal and zero crossing rate to detect the silent region. Taking both of them is important as only energy tends to miss some parts of sounds which are important.

B. Feature extraction component: The component generated feature vectors for the sound signals given to it. It generates Mel Frequency Cepstrum Coefficients and Normalised energy as the features that should be use to uniquely identify the given sound signal.

c. Recognition component: This is a Continuous, Multi-dimensional Hidden Markov Model base 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 component consists of Word based acoustic. Acoustic Model has a representation of how a word sounds. Recognition system makes use of this model while recognising the sound signal. The basic flow once the training is done, can be summaries as the sound input is taken from the sound recorder and is feeded to the feature extraction module. The feature extraction module generates feature vectors out of it which are then forward 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 [21].

The below given diagram is the just need of our work.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

Figure 3: Need of radio for real time communication

IV. METHODOLOGY

A. Formulation of the hypotheses: This idea combines two areas. One is real time communication for blind people and second is speech synthesis and recognition. This application will be made on Microsoft .Net. Microsoft provides the features of speech synthesis and speech recognition. Speech synthesis means text to speech conversion and speech recognition means speech to text conversion. Both ideas are combining. Now next task is to generate the sound of text understood. In this what is happening, there is a database which store the sound segments which is use accordingly. The effectiveness of text to speech conversion totally depends on how efficiently the sound segments are store in the database.
In speech to text conversion, firstly the audio signal is taken which includes strings, numbers and various pitches. This audio signal needs to check among the data store in the database. For better results recognizer should care about only the require stuff. So for this purpose application is a grammar which enhances the efficiency of the search. This idea is use to make an algorithm for searching the results of subjective questions from the database.

B. System software used

1) Visual Studio – 3.5

2) Sql Server – To create the knowledgebase

3) Speech SDK 5.1 for the conversion of

V. VOICE TO TEXT AND TEXT TO VOICE

Steps are as follows:

Blind user ask for information (to radio station) and the radio station replies for asked information e.g. hospital or Institute. The user selects the option by speech either for hospital by hospital or Institute, one of them.

This option acts as input data for machine.

⇩

The word is matched by speech recognition engine. It will make use of created grammars.

⇩

Check the input from database.

⇩

Result will be spoken by the system.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

The above scenario works for the completely speech enabled application and required features can be put in the application accordingly.

VI. SOURCES OF DATA

Data is provided by the user in the form of speech, and accordingly the response is given to user. Suppose user says one for hospital, then relative content is displayed to user in the form of text to speech conversion and the content is presented in audio form. That is input is provided by the system and converted into audio form.

A. Speech to text

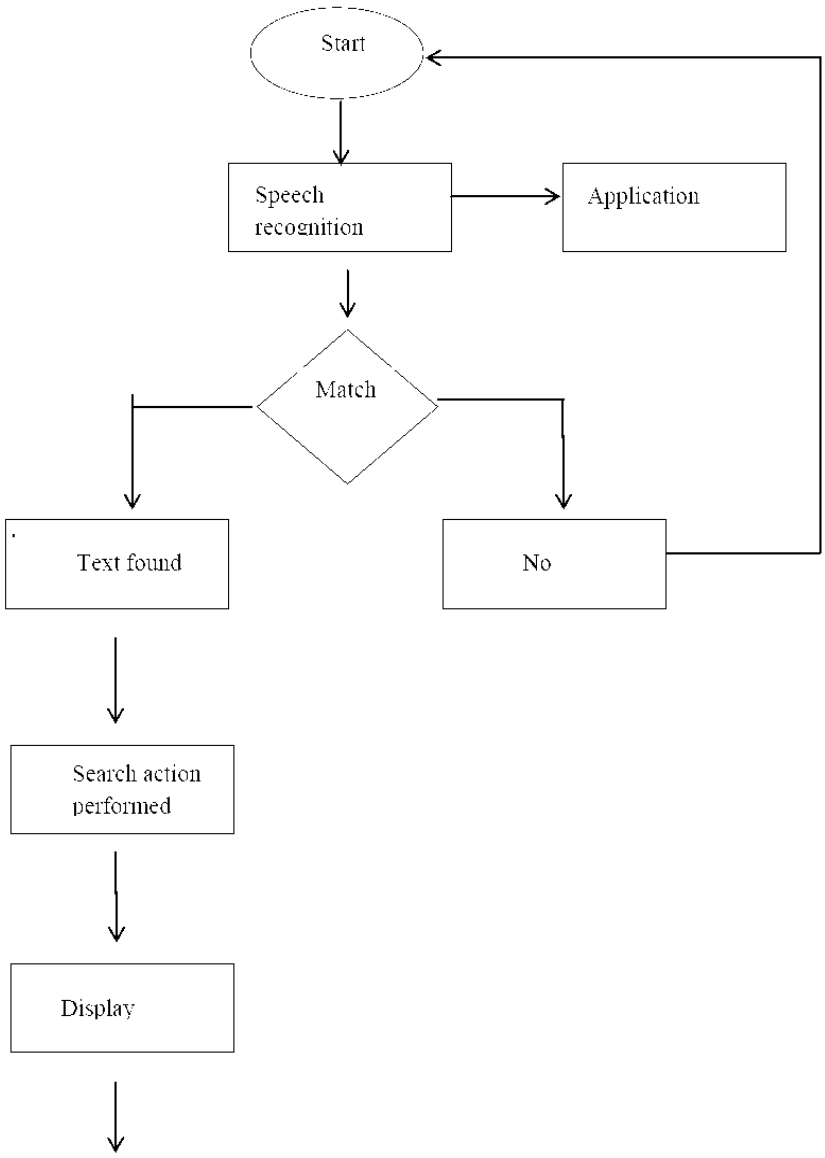

Figure 4: Block diagram of Speech to text

ISSN: 2319-8753
International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

B. Text to speech

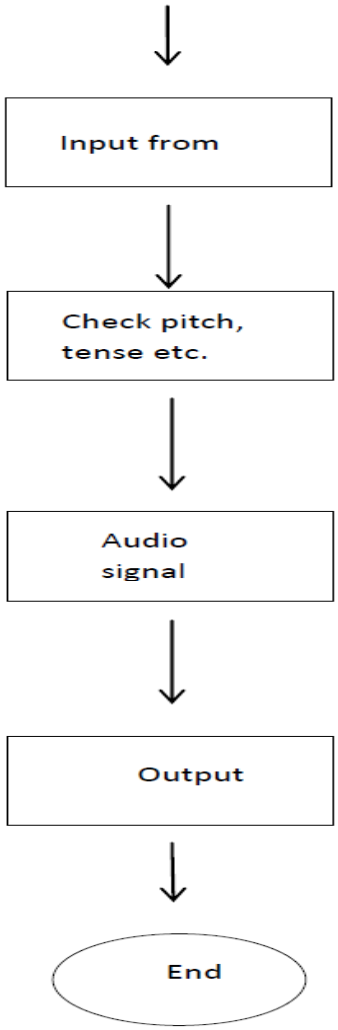

Figure 5: Block diagram for text to speech

VII. SPEECH SYNTHESIS ALGORITHM AND DECISION TREES

Decision tree inducers are algorithms that automatically construct a decision tree from a given dataset. Typically the goal is to find the optimal decision tree by minimizing the generalization error

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

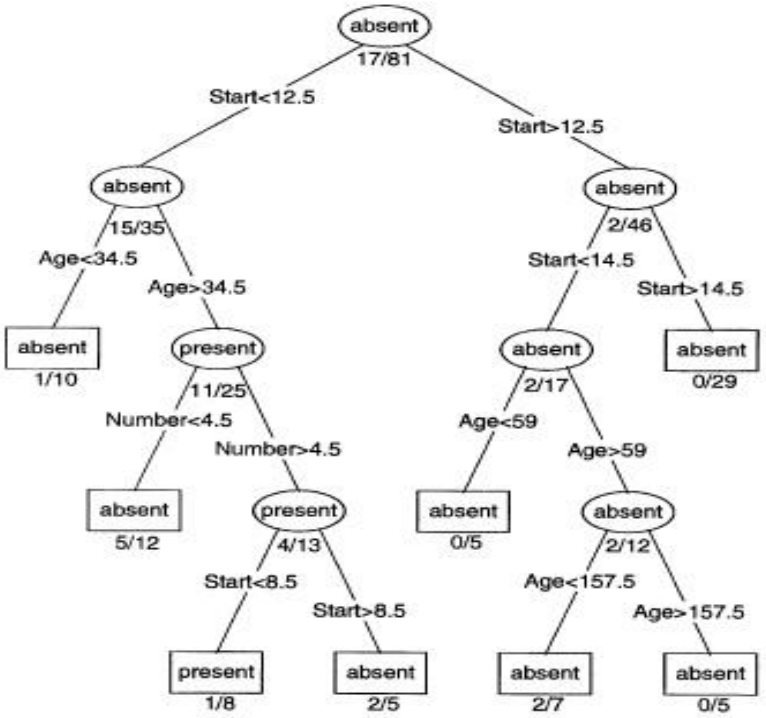

Figure 6: Block diagram for decision trees

A. Input from user array of voice signals {s1,s2,s3….sn}
B. If (input) {
signals sent to speech recognition engine
}
else {
//ask to provide input or time out
break;
}
C. check the speech grammars
D. If (true)
{ //search the content }
Else
{ // go to step1 }
E. Display output
F. Automatic conversion of text to speech
G. Pitch and tense checking done
H. Audio signal generation
I. Voice result will be provided to user

VIII. TOOLS OF DATA COLLECTION & ANALYSIS

Basically the idea is to convert the text into speech and vice versa. As we are not attaching any hardware into it hence we would start with a virtual call which the software shows of getting connected.

A. System provides options for you to speak hospitals, or Institute. It is done through speech synthesizer class object which is there in .NET framework 3.5 in an addition to speech.

B. When the user speaks anything, it is observed by the system through a speech recognizer and here the HMM algorithm is implemented with the use of the grammar builder class, a grammar is present there. The extracted and segmented word is matched with the grammar. If the match is found, there is a knowledge base from where the data will be fetched.

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

Suppose the user speaks one it means the user wants the information for the hospital. In recognizer class if it is matched with the grammar. Then the data will be fetched from the database accordingly.

C. Then again there is a reverse transformation of text to speech, again the HMM reverser would be applied and the user gets the information.
Before explaining steps required for speech recognition and synthesis lets discuss the reference required for this.
1. Microsoft speech object library needs to be used.
2. System, Speech class need to be referred.

Few header files which are required to perform speech recognition and speech synthesis.

i) Speech lib
ii) Speech recognition
iii) Speech synthesis

IX. EXPERIMENTAL RESULTS AND DISCUSSION

The user verification in non noisy area listed in given tables

Results of user verification in non noisy area			
Serial Number.	Word	Correct matching (no of users)	Wrong Matching (no of users)
1	Slow pitch	48	2
2	Normal pitch	50	0
3	Fast pitch	49	1
Total		147	3

Table 1: User verification in non noisy area

Positive Identification:
Total Number of user for each variable: 50
Variables: 3 (Slow, Normal and Fast)
The above result was obtained for 147 users reflect the correct matching out of 150.
In this case the results was found to be 98%.

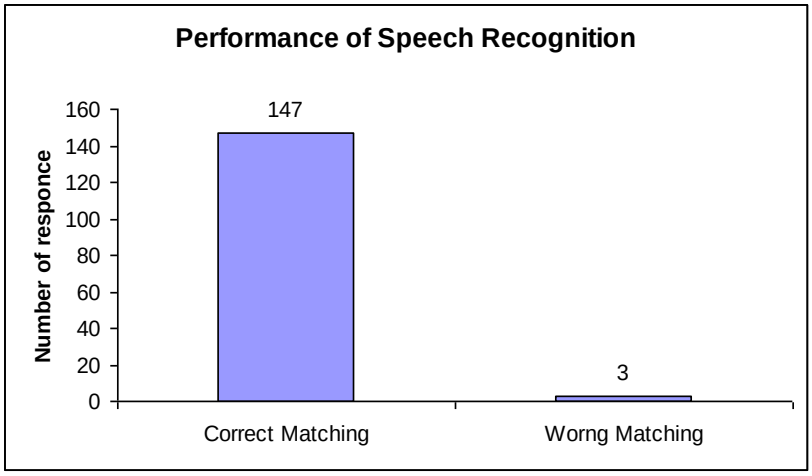

Figure 7: Block diagram for user verification

ISSN: 2319-8753

International Journal of Innovative Research in Science, Engineering and Technology
Vol. 2, Issue 2, February 2013

Results of user verification on the based of gender in non noisy area			
Serial Number.	Word (Normal Pitch)	Correct matching (no of users)	Wrong Matching (no of users)
1	Male	50	0
2	Female	50	0
3	Child	50	0
Total		150	0

Table 2: User verification (Different gender) in non noisy area

Positive Identification:
Total Number of male Users: 50
Total Number of female Users: 50
Total Number of child Users: 50

In the gender case the results was found to be 100%.

To test data we used the following shell script:

```
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 System.Speech.Recognition;
using System.Speech.Synthesis;

namespace voice_calculator
{
    public partial class Form1 : Form
    {
        SpeechRecognizer speechreco = new SpeechRecognizer();
        SpeechSynthesizer sp_sz = new SpeechSynthesizer();
        long a = 0;
        long b = 0;
        long res = 0;
        String opr = "";
```

X. CONCLUSION AND FUTURE WORKS:

This proposed model enables the visually impaired person to function independently like normal humans. It improved the quality of their life. The problem with the existing system is the delay encountered and lack of efficiency. The main aim of serving the mankind is achieved by providing sense of sight to the needy by this paper. In recent there is no much time for user to achieve the information due to busy life style. So in these days the health and career problems are main things to the youths. We provided a comprehensive solution of real time information communication between user and machine on same time when needed by users. In future we want to do is:

A. Evaluate the learner by means of subjective questions as well with the help of radio.

B. User will give the input to the system. This input will be checked against the database used for the questionnaire. On the basis of most appropriate matching results will be displayed in the sound and this sound broadcast with the help of radio station to the listener.

C. To minimize the noisy error rate.

XI. ACKNOWLEDGEMENT

I would like to express my deepest gratitude to Navdeep Kaur (Dissertation Mentor) for her valuable guidance. It is only with her guidance that I could take up initiative of such a good topic of thesis. I am also very thankful to Mr. D. R Sharma, All India Radio Chandigarh for giving me opportunity to propose and implement my work.

REFERENCES

[1] V.D. Kanna and S.Aswin Amirtharaj, P.V.Deepak Govind Prasad, N.Sriram Prabhu *"Design of a FPGA based Virtual Eye for the Blind"* 2011 2nd International Conference on Environmental Science and Technology IPCBEE vol.6 (2011 IACSIT Press, Singapore
[2] B. H. Juang and F. K. Soong, *"Hands-free telecommunications,"* Proc. International Conference on Hands-Free Speech Communication, pp. 5–10, 2001.
[3] J. F. Cardoso, *"Eigenstructure of the 4th-order cumulant tensor with application to the blind source separation problem,"* Proc. ICASSP'89, pp. 2109–2112, 1989.
[4] C. Jutten and J. Herault, *"Blind separation of sources part i: An adaptive algorithm based on neuromimetic architecture,"* Signal Processing, vol. 24, pp. 1–10, 1991.
[5] P. Comon, "*Independent component analysis, a new concept?"* Signal Processing,vol. 36, pp. 287–314, 1994.
[6] A. J. Bell and T. J. Sejnowski, *"An information-maximization approach to blind separation and blind deconvolution,"* Neural Computation, vol. 7, no. 6, pp. 1129–1159, 1995.
[7] S. Amari, S. Douglas, A. Cichocki, and H. Yang, *"Multichannel blind deconvolution and equalization using the natural gradient,"* Proceedings IEEE International Workshop on Wireless Communication, pp. 101–104, 1997.
[8] P. Smaragdis, *"Blind separation of convolved mixtures in the frequency domain,"* Neurocomputing, vol. 22, no. 1-3, pp. 21–34, 1998.
[9] S. Choi, S. Amari, A. Cichocki, and R. Liu, *"Natural gradient learning with a nonholonomic constraint for blind deconvolution of multiple channels,"* Proceedings of International Workshop on ICA and BSS, pp. 371–376, 1999.
[10] N. Murata and S. Ikeda, *"An on-line algorithm for blind source separation on speech signals,"* Proc. of NOLTA98, vol. 3, pp. 923–926, 1998.
[11] T. Takatani, T. Nishikawa, H. Saruwatari, and K. Shikano, *"Highfidelity blind separation of acoustic signals using simo-model-based ica with information geometric learning,"* Proc. IWAENC2003, pp. 251–254, 2003.
[12] T. Nishikawa, H. Saruwatari, and K. Shikano, *"Stable learning algorithm forblind separation of temporally correlated acoustic signals combining multistage ica and linear prediction,"* IEICE Trans. Fundamentals, vol. E86-A,no. 8, pp. 2028–2036, 2003.
[13] L. Parra and C. Spence, *"Convolutive blind separation of non-stationarysources,"* IEEE Transactions on Speech and Audio Processing, vol. 8, pp.320–327, 2000.
[14] H. Saruwatari, T. Kawamura, T. Nishikawa, and K. Shikano, *"Fastconvergence algorithm for blind source separation based on array signal processing,"* IEICE Trans. Fundamentals, vol. E86-A, no. 4, pp. 286–291,2003.45
[15] T. Nishikawa, H. Saruwatari, and K. Shikano, *"Blind source separation of acoustic signals based on multistage ica combining frequency-domain ica and time-domain ica,"* IEICE Trans. Fundamentals, vol. E86-A, no. 4, pp. 846–858, 2003.
[16] H. Sawada, R. Mukai, S. Araki, and S. Makino, *"Polar coordinate based nonlinear function for frequency domain blind source separation,"* IEICE Trans. Fundamentals, vol. E86-A, no. 3, pp. 590–596, 2003.
[17] S. Araki, S. Makino, Y. Hinamoto, R. Mukai, T. Nishikawa, and H. Saruwatari, *"Equivalence between frequency domain blind source separation and frequency domain adaptive beamforming for convolutive mixtures,"* EURASIP Journal on Applied Signal Processing, vol. 2003, no. 11, pp. 1157–1166, 2003.
[18] Y. Mori, H. Saruwatari, T. Takatani, S. Ukai, K. Shikano, T. Hiekata,Y. Ikeda, H. Hashimoto, and T. Morita, *"Blind separation of acoustic signalscombining simo-model-based independent component analysis and binary masking,"* EURASIP Journal on Applied Signal Processing, vol. 2006, pp. Article ID 34 970, 17 pages, 2006.
[19] M.A.Anusuya, S.K.Katti *Sri Jaya chamarajendra College of Engineering Mysore,* India International Journal of Computer Science and Information Security, Vol. 6, No. 3, 2009
[20] Dat Tat Tran, *"Fuzzy Approaches to Speech and Speaker Recognition ,"* A thesis submitted for the degree of Doctor of Philosophy of the university of Canberra.
[21] Simon Kinga and Joe Frankel, Recognition ,*Speech production knowledge in automatic speech recognition* , Journal of Acoustic Society of America,2006.
[22] Robert Batusek and Ivan Kopecek, *"User Interfaces for the Visually Impaired people",* Masaryk University, 2000.

BIOGRAPHY

Mrs Ranu Dixit is a Research Scholar at Chandigarh Engineering College, Landran Mohali Punjab 140307 India. Her research interests are in the field of Information Technology for blind peoples. Her specialization is Speech recognition. She completed Polytechnic diploma in Information Technology, BE in Information Technology and presently pursuing M. Tech in Information Technology.