

INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING & TECHNOLOGY (IJCET)

ISSN 0976 – 6367(Print)

ISSN 0976 – 6375(Online)

Volume 4, Issue 2, March – April (2013), pp. 425-436

© IAEME: www.iaeme.com/ijcet.asp

Journal Impact Factor (2013): 6.1302 (Calculated by GIS)

www.jifactor.com



.....

HINDI SPEECH ENABLED WINDOWS APPLICATION USING MICROSOFT SAPI

Kamlesh Sharma¹, Dr. T.V. Prasad², Dr. S. V. A. V. Prasad³

¹(Research Scholar, Dept. of CSE, Lingaya`s University , Faridabad, Haryana, India)

²(Dean of computer Science, Visvodaya Technical Academy, Kavli, Andhra Pradesh, India)

³(Dean of R&D, Lingaya`s University , Faridabad, Haryana, India)

ABSTRACT

People with disability such as visual impaired, elderly for whom it's very hard to identify the screen text and area where the keyboard and mouse may not be an appropriate means of communication between system. So it would be a real relief to have the option to use ears to listen to the contents and use voices to navigate and control the computer systems. Even if sometimes it is for normal people, it would be more comfortable to work with speech enabled applications. Microsoft has designed an interface called SAPI (Speech Application Programming Interface) which supports dynamic speech input and output, and is integrated in our current operating systems. With the API it is possible to develop speech enabled applications without caring about the details of synthesis and recognition. In this paper, a Hindi Speech enabled Windows Application (HSeA) is presented to demonstrate the use of speech-enabled application using Microsoft SAPI in Microsoft Windows Operating Systems.

Keywords : HSeA, Operating System, speech enabled, Windows application, SAPI.

I. INTRODUCTION

An operating system is software that manages all the resources of a computer, hardware and software, a number of applications and provides an environment in which a user can execute programs in a convenient and efficient manner [1]. However, the principles and concepts used in the operating systems were not standardized in a day. In fact, the operating systems have been evolving through the years [2][3]. In the operating systems, every program and application required full interaction of mouse and keyboard.

Computers are no longer a field of education for the Educational Institutes. They have changed from a large desktop computer or laptop to small pocket type PC like tablet and smart phone. In developed countries, using a computer has become the basic requirement like reading and writing a book. According to the survey done by Asia's premier monthly magazine on ICT in education, 85% of the total people on the earth will be using the Smartphone's or mobile and computer by 2020. The importance of digital information has already reached such an enormous level that many companies, corporate sector and politicians around the world try to find a way to paperless e-governance. But a question arises, is this really accessible to everybody?

The growing need of a computer system and the application programs eventually made speech enabled operating systems a necessity. Most of the software coded today is only accessible through mouse and keyboard. But the expected improvements to the SAPI version included in Windows Vista may lead to a wave of new English speech enabled applications [4][5]. Microsoft Windows is compatible with a wide variety of assistive technology products such as screen readers, magnifiers, and specialty hardware that meet the needs of computer users with all types of physical impairments. Full integration for speech synthesis and recognition as well as support for native and managed code could be part of the Windows operating system [6].

Speech is the most natural way of communication. It also provides an efficient means of man-machine communication. Generally, transfer of information between human and machine is accomplished via keyboard, mouse etc. But humans can speak more quickly instead of typing. Speech interfacing provides the ways to these issues [7]. Speech interfacing involves speech synthesis and speech recognition.

Speech Recognition is a technology that allows the computer to identify and understand words spoken by a person using a microphone. Speech recognition allows a computer to interpret any sound input (through either a microphone or audio file) to be transcribed or used to interact with the computer. Whereas, Speech synthesizer takes the text as input and converts it into the speech output i.e. it acts as text to speech converter. Speech recognizer converts the spoken word into text [4].

In this paper, an attempt has been made to develop a Hindi Speech Recognition and Synthesis application as an assistive technology to provide a solution for the Hindi speaking people. The system is designed in Microsoft .NET framework using C# Programming in Microsoft Visual Studio 2008 Environment. Microsoft Windows Speech Application Programming Interface (SAPI) 5.3 and, system speech recognition and system speech synthesis namespaces are used for speech to text conversion and vice-versa.

II. NEED OF SIGNIFICANCE

Speech recognition has not only helped the users to access information and knowledge but also entertainment by simply voicing their needs. This system promises to be of significant advantages in this area where a keyboard and mouse may not be an appropriate mean for communication between system and user and natural communication is desired. This system includes speech to control application at the same time hands and eyes may be busy. Additionally, such system can be widely applied for people who have vision related disability, impairment of motor control, etc.

The goal of this research is to incorporate Hindi speech recognition (speech-to-text) and Hindi speech synthesis (text-to-speech) technology [8] into the applications and providing the solution for Hindi speaking people. This technology does not require any additional equipment besides computers and Headphone with mike facility[9][11].

III. METHODOLOGY

The methodology for the current system is as follows in the figure. 1

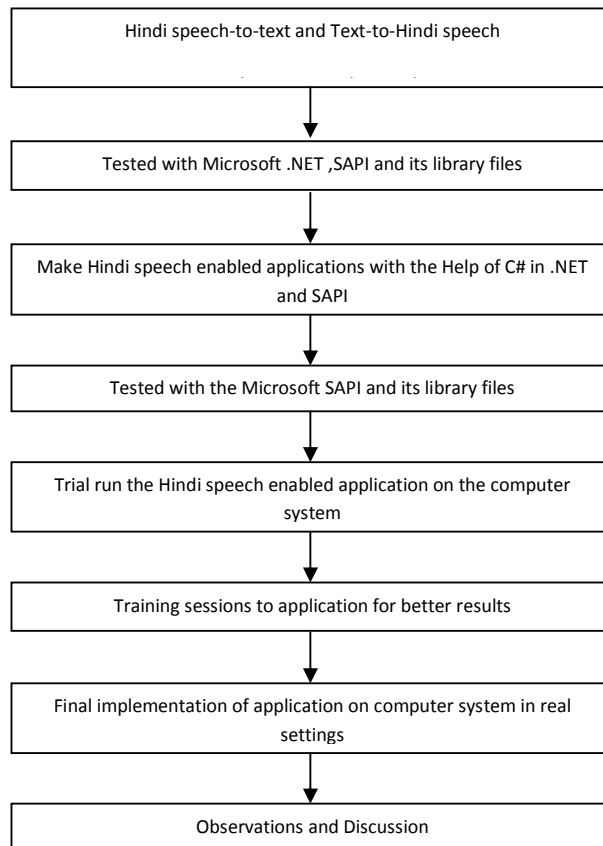


Fig. 1: Methodology of work

IV. WINDOWS APPLICATION DESIGN

The manner in which users interact with a program or an application is known as its user interface. The user interface controls how data is entered and how information is displayed. The primary aim of Hindi speech enabled applications is to improve interaction between user and machine. For this purpose applications are developed in the .NET Framework using C# , SAPI and uses Microsoft SQL Server.

4.1 Fundamental Basics

4.1.1 The Microsoft .NET Framework is a software technology that is available with several Microsoft Windows operating systems. It includes a large library of pre-coded solutions to common programming problems, a runtime or virtual machine that manages the execution of programs written specifically for the framework, and a set of tools for developing and configuring the applications. The .NET Framework is a key Microsoft offering and is intended to be used by most new applications created for the Windows platform. The .NET Framework has two main components: the common language runtime and the .NET Framework class library. The .NET Framework is an environment for building, deploying, and running Web Services and other applications. It consists of three main parts: The Common Language Runtime, The Framework classes, and The ASP.NET. in NET is a general purpose software development platform, similar to Java [12].

4.1.2 Microsoft Visual C# is Microsoft's implementation of the C# specification, included in the Microsoft Visual Studio suite of products. It is based on the ECMA/ISO specification of the C# language, which Microsoft also created. While multiple implementations of the specification exist, Visual C# is by far the one most commonly used. Visual C# is also heavily used by ASP.NET web sites and stand alone applications based on the .NET Framework. C# is a new programming language, very similar to Java. An extensive class library is included, featuring all the functionality one might expect from a contemporary development platform - windows GUI development (Windows Forms), database access (ADO.NET), web development (ASP.NET), web services, XML etc [13].

4.1.3 Microsoft Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop console and graphical user interface applications along with Windows Forms applications, web sites, web applications, and web services in both native code together with managed code for all platforms supported by Microsoft Windows, Windows Mobile Windows CE, .NET Framework, .NET Compact Framework and Microsoft Silverlight [14].

4.1.4 Microsoft Speech Application Programming Interface (SAPI) is an API developed by Microsoft to allow the use of speech recognition and speech synthesis within Windows applications. It is possible for a 3rd-party company to produce their own Speech Recognition and Text-To-Speech engines or adapt existing engines to work with SAPI [14]. SAPI 5 however was a completely new interface, released in 2000.

4.1.5 The System.Speech.Synthesis namespace can be used to access the SAPI synthesizer engine to render text into speech using an installed voice, such as Microsoft Anna. The SAPI 5.3 synthesizer now supports the W3C standard Speech Synthesis Markup Language (SSML), a markup language that allows you to finely tune how the synthesizer will produce words, such as pronunciation, speed, volume and pitch, of the produced phrase [15].

4.1.6 The System.Speech.Recognition engine is used to recognize a user's voice and convert it into text. The SAPI 5.3 recognition engine now supports the W3C standard; Speech Recognition Grammar Specification (SRGS), a markup language that defines how and what words are recognized. SAPI 5.3 also added support for Semantic Interpretation [15].

The block diagram of Hindi speech enabled applications are shown in fig. 2. Where it is shown how the user and the system interacts with each other using the Hindi enabled application (HSeA). A good quality headphone is required for generating an audio signal, which then amplified and digitized, then finally transmitted it to the interface in which user query is received. The interface running on laptop computer analyzed the signal and then send a request to the computer system. The computer performs the action and send the result back to the interface. The voices utterances are of words fed to statistical speech recognition model using SAPI where the word that were uttered most likely are determined. A database was constructed with a list of words defining specific subject like mail, calculation, document etc [9].

The uttered words are compared to the database words, if uttered word match is found, the prompted application actuated and if a match is not found the user is prompted for the error and the recorder is again initialized.

4.2 Different Hindi speech enabled application(HSeA) are as follows

- Hindi speech enabled Word.
- Hindi speech enabled Calculator.
- Hindi speech enabled outlook.

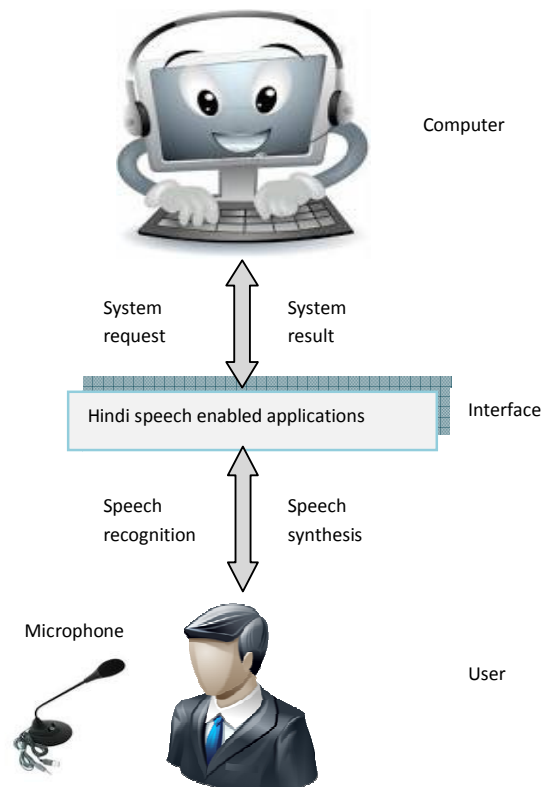


Fig. 2. Interaction between users, HSeA

4.2.1 Hindi speech enabled word:- This is one of the application which is enabled by Hindi speech as shown in fig 2. In this application user control word by speaking to it a microphone, entering text , or issuing commands to the computer, e.g. to open a new document, to save the document, to print a document etc. User can talk to our computer using a set of pre-defined commands and instruction and computer will respond in the same way. For example user can say : “dustavej bana”, and the computer would respond: “open new document” . or user can say : “dustavej surkshit karna” and the computer will save the document.

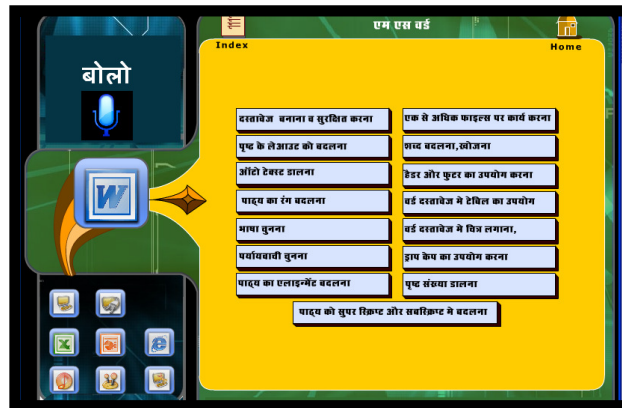


Fig 3. Hindi speech enabled word

4.2.2 Hindi speech enabled calculator:- It is an application which controls the calculator by Hindi speech as shown in fig 3. It is used for mathematical calculation like jama, ghata, guna, bhag. It takes numerical values as input and performs the operation. For example to add “3+5=8”, It takes speech input “teen jama panch” and shows the result on speech “lao” then “8” displayed in small textbox.

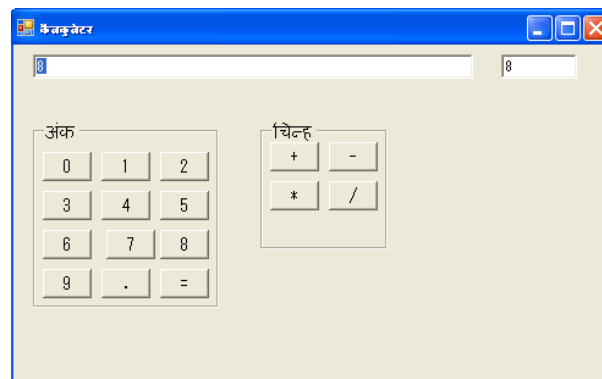


Fig 3 Hindi speech enabled calculator

4.2.3 Hindi speech enabled outlook:- This application is used for sending and reading mail through Hindi speech as shown in fig 4. It consists two legend one for sending mails and another one for reading mails. The sending mail legend is named as “pado”. It consists two text boxes which are label as “kisko” or “sandesh”. The “kisko” textbox is used for email addresses and “sandesh” for the message. When the user speak kisko the cursor activates in

the textbox and will waiting for an email address after then the cursor move to sandesh textbox on speech “sandesh”. The sandesh textbox is used for the content of the mail. User dictate the content and send the mail to lebeled user’s. The “Pado” legend is used to read the mails. When user raise a command “pado”. It fetches the mail from the mail server an shows then on the system.

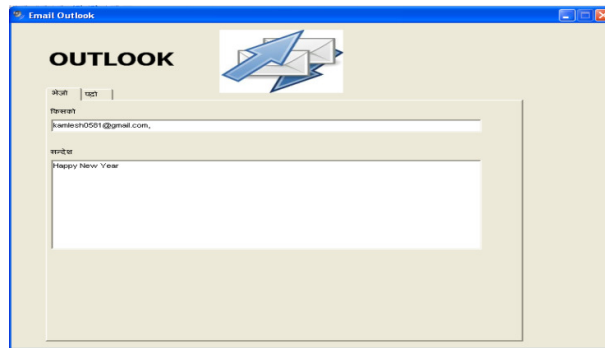


Fig 4. Hindi speech enabled outlook

V. IMPLEMENTATION

The applications can be installed on desktop or laptop having the following system specifications:

- Computer System Configuration: Microsoft Window XP and above, 512 MB RAM, 40 GB, HD, Realtek High.
- Definition Audio Sound Card.
- Softwares Required: Microsoft Visual Studio 2008,
- Microsoft .NET 3.5 Framework, Microsoft SAPI.
- External peripheral required: A noise cancelling Headphone with mice facility.

5.1 Training:- Training sessions were conducted with Microsoft Speech Recognition and Speech Synthesis engines for effective and efficient Implementation of Hindi speech enabled system in a noise free environment. Samples were conducted by the college student and all the students were familiar with Hindi language. Some samples were also conducted from deaf and dumb school. The system was trained with randomly selected students. Initially limited commands were given to the student for the training and testing purpose. Then the student was given extensive training for speech recognition. Feedback and error message were taken from each of the students .

5.2 Implementation of HSeA

The implementation of HSeA for windows application is explained by following step:-

- Computer or laptop which is installed with Hindi enabled speech applications given to the users.
- The noise free environment was provided for effective communication.
- A list of commands is handed to the user.

- By using Speech Application Programming Interface (SAPI) the input commands match with the stored words to obtain a perfect match between the applied command and the stored command.
- Once such matching is accomplished, the particular matched word is converted into the system command and execute the voice command. The fig 5. shows the working of HSeA.

VI. TESTING AND RESULTS

Experiments were conducted to evaluate the baseline system and the improved HSeA. The SAPI was used in these experiments. A set of 20 commands was chosen from the set of 100 training commands to test the horse. A set of 20 students was randomly chosen from 60 students) and were asked to raise a command to the HSeA to retrieve the result accordingly and show the user. The system also tests on unknown user which were not trained.

The success of each trial was based on whether the system was able to retrieve the required information to the user or not. For example user will execute the commands to the HSeA and system respond accordingly. The table 1. Shows the commands probability testing.

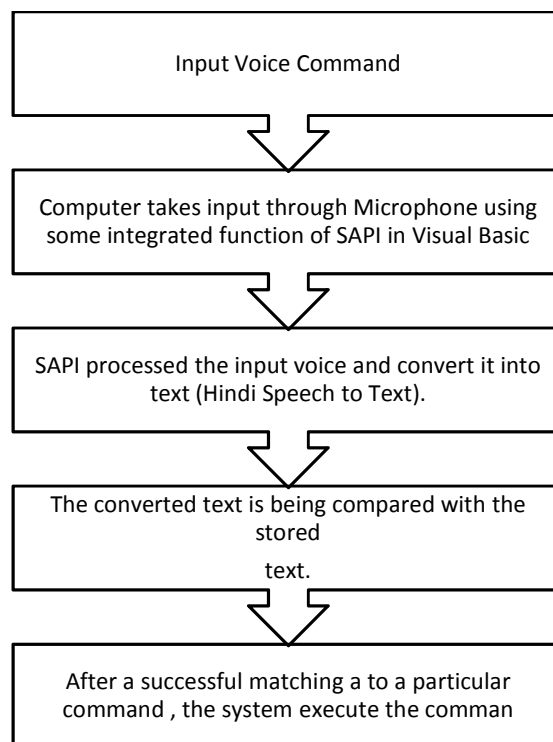
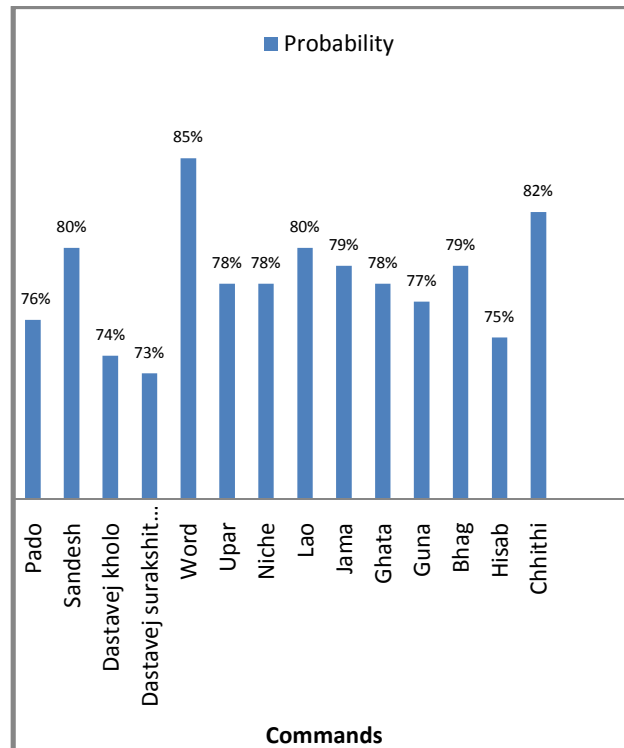


Fig 5. Working of HSeA

Command	Testing	Probability
Pado	20	76%
Sandesh	20	80%
Dastavej kholo	20	74%
Dastavej surakshit karo	20	73%
Word	20	85%
Upar	20	78%
Niche	20	78%
Lao	20	80%
Jama	20	79%
Ghata	20	78%
Guna	20	77%
Bhag	20	79%
Hisab	20	75%
Chhithi	20	82%

Table 1. Commands probability testing



Graph 1. Commands probability Testing

VII. CONCLUSION AND FUTURE RESULTS

This paper presents a scheme proposed to control computer systems through Hindi voice of different users. The results of this experiment suggest that native Indians who are not able to use the computer and/or lack English skills [10] will be able to use Hindi voice based control. 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. Today, while considering this question, and uses where these technologies will be needed and desire, which would warrant R&D expenditures. There are a number of scenarios where speech recognition is either being delivered, developed for, researched or seriously discussed like computer and video games, precision surgery, domestic applications, wearable computers etc. [11]

There are several challenges the system will need to deal with in the future. First, the overall robustness of the system must be improved to facilitate implementation in real life applications involving telephone and computer systems. Second, the system must be able to reject irrelevant speech that does not contain valid words or commands. Third, the recognition process must be developed so that commands can be set in continuous speech. And finally, the voice systems must be able to become viable on low-cost processors. Thus, this will enable the technology to be applied in almost any product.

REFERENCES

- [1] Silberschatz, A., Galvin, P.B. and Gagne, G., Operating System Principles. 7th ed., John Wiley & Sons, 2006.
- [2] Tanenbaum, A.S. and Woodhull, A.S., Operating Systems Design and Implementation. 3rd ed., Prentice-Hall, 2006.
- [3] Alan C. Bomberger, A. Peri Frantz, William S. Frantz, Ann C. Hardy, Norman Hardy, Charles R. Landau, Jonathan S. Shapiro, The keykos® Nanokernel Architecture in Proceedings of the USENIX Workshop on Micro-Kernels and Other Kernel Architectures, USENIX Association, April 1992. Pp 95-112.
- [4] Microsoft speech SDK 5.1, <http://www.microsoft.com/downloads/details.aspx?FamilyID=5e86ec97-40a7-453f-b0ee->
- [5] C# Corner 2004, 'Speech Recognition using C#', <http://www.csharpcorner.com/uploadfile/ssrinivas/speechrecognitionu>
- [6] Takahiro Ikeda, Shin-ya Ishikawa, Kiyokazu Miki, Fumihiro Adachi, Ryosuke Isotani, Kenji Satoh and Akitoshi Okumura, Speech-Activated Text Retrieval System for Cellular Phones with Web Browsing Capability, Proceedings of PACLIC 19, the 19th Asia-Pacific Conference on Language, Information and Computation.
- [7] Michael D. Goller and Stuart E Goller, Speech Interface for search Engine". United state patent, Jun. 22 2010, shett no 1 to 4.

- [8] Kamlesh Sharma* and Dr. T. V. Prasad , A Text-T0-Hindi Speech Interface for Web Browsing, National Conference on Advances in Knowledge Management (NCAKM'10) at Lingaya's University 2010.
- [9] Kamlesh Sharma, Dr. S. V. A. V. Prasad, Dr. T. V. Prasad, A Hindi Speech Actuated Computer Interface for Web Search, International Journal of Advanced Computer Science Applications, Vol. 3, No. 10, ISSN : 2156-5570(Online), 2158-107X (Print), Impact Factor: 1.187, arXiv:1211.2741, 2012.
- [10] Kamlesh Sharma, T. Suryakanthi, Dr. T. V. Prasad, Exploration of Speech enabled System for English, Proc. of the International Conference on System Modeling and Advancement in Research Trends (SMART), Teerthankar Mahaveer University, Moradabad, UP, India, 2012.
- [11] Kamlesh Sharma, Dr. T. V. Prasad, "CONATION" : English Command Input/Output System for Computer , Proc. of the International Conference on Science, Engineering & Spirituality (ICES'10) at S.E.S. College of Engineering, Navalnagar jointly with IEEE Bombay Section and IEEE Computer Society, 2010.
- [12] F. Reena Sharma and S. Geetanjali Wasson, Speech Recognition and Synthesis Tool: Assistive Technology for Physically Disabled Persons, Proc. Of the International Journal of Computer Science and Telecommunications [Volume 3, Issue 4, April 2012.
- [13] .NET Framework Conceptual Overview, Microsoft Developer Network Platform, Retrieved on March 10, 2012 from <http://msdn.microsoft.com/enus/library/w0x726c2%28v=vs.90%29.aspx>
- [14] 'Getting Started with C#', Microsoft Developer Network Platform, Retrieved on March 10, 2012 from <http://msdn.microsoft.com/enus/library/z1zx9t92.aspx>
- [15] Dunn, Michael (2007) Give Applications a Voice, Speech synthesis and recognition in .NET, Tech Brief Articles, Retrived on 14th March, 2012 http://reddevnews.com/articles/2007/02/15/give-applications-avoice.aspx?sc_lang=en
- [16] Debashis Chakraborty, Sutirtha Ghosh and Joydeep Mukherjee, "Efficient Text Compression using Special Character Replacement and Space Removal", International Journal of Computer Engineering & Technology (IJCET), Volume 1, Issue 2, 2010, pp. 38 - 46, ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375
- [17] Gunjan Singh, Avinash Pokhriyal and Sushma Lehri, "Fuzzy Rule Based Classification and Recognition of Handwritten Hindi Curve Script", International Journal of Computer Engineering & Technology (IJCET), Volume 4, Issue 1, 2013, pp. 337 - 357, ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375
- [18] Prof. S.A.Ubale and Dr. S.S. Apte, "Study and Implementation of Code Access Security with .Net Framework for Windows Operating System", International Journal of Computer Engineering & Technology (IJCET), Volume 3, Issue 3, 2012, pp. 426 - 434, ISSN Print: 0976 – 6367, ISSN Online: 0976 – 6375

ABOUT AUTHORS



Ms. Kamlesh Sharmas received her masters in Computer Sc. & Engg. degree from Maharshi Dayanand University, Rohtak, India in 2009. She is currently associated with at Lingaya's University, Faridabad in the Dept. of Comp. Sc. & Engg. as Research Scholar. She has over 7 years of teaching experience at under graduate and graduate levels. Her areas of interest are artificial intelligence, operating systems, web mining, Database Management Systems, etc.



Dr. T. V. Prasad has over 17 years of experience in industry and academics. He received his graduate and master's degree in Computer Science from Nagarjuna University, AP, India. He was with the Bureau of Indian Standards, New Delhi for 11 years as Scientist/Deputy Director. He earned PhD from Jamia Millia Islamia University, New Delhi in the area of computer sciences/ bioinformatics. He has worked as Head of the Department of Computer Science & Engineering, Dean of R&D and Industrial Consultancy and then as Dean of Academic Affairs at Lingaya's University, Faridabad. He is with Visvodaya Technical Academy, Kavali as Dean of Computing Sciences. He has lectured at various international and national forums on subjects related to computing. Prof. Prasad is a member of IEEE, IAENG, Computer Society of India (CSI), Indian Society of Remote Sensing (ISRS) and APBioNet. His research interests include bioinformatics, artificial intelligence (natural language processing, swarm intelligence, robotics, BCI, knowledge representation and retrieval). He has over 75 papers in different journals and conferences, and also has six books and two chapters to his credit.



Dr. S. V. A. V. Prasad has over 30 years of experience in industry and academics. He has received his master's degree in Electronics & Communications Engg. from Andhra University, AP, India. He earned PhD from Andhra University, Waltair, Visakhapatnam, India. He was with leading research and manufacturing companies in New Delhi, India. He also taught at leading institutions like the Delhi College of Engg. (now Delhi Technological University), Delhi for many years.. He has worked as Head of the Department of Electronics & Communications Engg., Dean of Academic Affairs and as Dean of R&D and Industrial Consultancy at Lingaya's University, Faridabad. He has lectured at various forums on subjects related to electronics, communications, audio engineering, signal processing, etc. Prof. Prasad is a member of IEEE, ISTE, etc. His research interests include audio engineering, signal processing, etc.. He has large number of papers in different journals and conferences.