Operating PC using Regional Language Speech on Single GUI

Dr. S. Jayanthi[1] , S. Vamshi Krushna[2],
Professor/Dean(R&D)[1], Assistant Professor[2],
Samskruti College of Engineering and Technology
Kondapur, Telangana , India
vamshilike@gmail.com[1], nigilakash@gmail.com[2]

Abstract

Some decades back, computer and its operations remained mystery to many, but now it is not so. In today's world, almost all operations are performed in an easy way with the aid of computers and also in a highly interactive user friendly manner. Recent trends like touch screen have made human-computer interaction easy, other than this application specific software such as speech to text converters (which also makes HCI easier) are available. ***Speech recognition (SR) systems*** *allow people to control computer by speaking to it through a microphone, either entering text, or issuing commands to the computer. Application systems that utilize recognition technologies, such as speech recognition, provide human-machine interface that could aid people more easily in operating system device or help those who are physically unable to interact with computers through traditional input devices such as mouse or keyboard. As we have seen, speech recognition technology is widely used between the device interfaces and human.*

Keywords: SR systems, Discrete & Continuous Speech, SAPI, Navigation Path, RSI, Dyslexia.

1. Introduction

SR systems have been around for over twenty years, but the early systems were very expensive and required powerful computers to run. The technology behind speech output is of two types. **Discrete speech**, i.e. the user had to speak one word at a time, with a short pause between words. **Continuous speech**, allowing the user to speak in a more natural way.

In this paper work, we provide the proposed software design 'operating pc using speech' is used operating pc directly without need of mouse and keyboard. This software improves the system performance and takes less time operate the system and it is secure and more efficient. We provide a generic and visual interfacing framework for bridging the interface between application systems and recognizers through the application system's front end, applying a visual level interfacing without requiring the detailed system design and programming knowledge.

When end user speaks a command through interface, the interface would analyze and recognize it, and then interact directly with application systems through calling the API functions. The proposed system could be applied to GUI based commercial software without accessing their internal code.

When installing and executing our proposed interfacing framework under all windows environment, user can interact with computer. Finally, we apply some examples to demonstrate the applicability and feasibility of the proposed works.**US/UK English slang problem** will be tried to overcome by implementing the system in Indian Languages.

1.1 Speech API

The Microsoft Speech API better known as Microsoft SAPI provides a high-level interface between an application and speech engines. It dramatically reduces the code overhead required for an application to use speech recognition and text-to-speech, making speech technology more accessible and robust for a wide range of applications. Microsoft Windows XP and Windows Server 2003 include SAPI version 5.1. Windows Vista and Windows Server 2008 include SAPI version 5.3, while Windows 7 includes SAPI version 5.4.SAPI implements all the low-level details needed to control

and manage the real-time operations of various speech engines.

The two basic types of SAPI engines are text-to-speech (TTS) systems and speech recognizers. TTS systems synthesize text strings and files into spoken audio using synthetic voices. Speech recognizers convert human spoken audio into readable text strings and files.

1.2 How to Speak

Continuous speech is not quite the same as natural speech! It is important to speak clearly, without slurring words, otherwise, a phrase like "the stuff he knows" might be interpreted as "the stuffy nose". Pairs of small words like "the" "and" "a", or "that" and "but" are most likely to be misrecognised, particularly in a phrase like "with the book".

It is important to enunciate each word clearly, rather than running them together as people tend to do in natural speech. The user should try to be aware of words that are being mis-recognized. It is possible to give the system additional training in the pronunciation of individual words and it may be that the user might want to slightly alter their pronunciation of some short words if they are consistently mis-recognized.

It is important to try to stay relaxed when using a SR system, particularly when things start to go wrong. It is very easy to become tense and frustrated when the system fails to recognize words. The inevitable change of voice only leads to a further decrease in recognition rates.

2. Technologies Involved

The technologies behind speech are:

1**. Discrete speech**, i.e. the user had to speak one word at a time; with a short pause between words. Discrete speech systems were particularly useful for people who had difficulty in forming complete phrases in one utterance and were generally more tolerant of non-standard speech. The focus on one word at a time was also useful for people with a learning difficulty. They did, however, force people to speak in an un-natural manner, which could cause voice strain.

2. Continuous speech, over the past few years the field of speech recognition has become dominated by the use of continuous speech systems, allowing the user to speak in a more natural way. Continuous speech is not quite the same as natural speech! It is important to speak clearly, without slurring words; otherwise, a phrase like "the stuff he knows" might be interpreted as "the stuffy nose".

3. Background Noise

Background noise can be a problem with speech recognition systems, causing words to be misrecognised, or phantom words to appear in your text. It is best to use the system in a quiet room – a typical classroom is not necessarily a good location!

As software and hardware has improved over the past few years, it has become less of a problem. If the user does experience a problem, then the use of a good microphone, e.g. the TalkMic or Andrea NC61, both of which have built-in technology to cancel out background noise, should reduce the problem.

4. Who Can Benefit From The Use Of Speech Recognition Systems?

People with Physical Access Difficulties

People with conditions such as RSI (Repetitive Strain Injury), arthritis, high spinal injury can usually benefit from SR systems, particularly where they have past experience of using computers. The choice of system will depend on the level of disability and the nature of the tasks involved. **People with**

Specific Learning Difficulties (Dyslexia) can also be benefited

5. What Factors can Influence Success?

5.1 Speech Consistency

Consistency of speech is much more important than voice quality. Many people with quite dysarthric speech are able to use discrete

speech recognition systems provided that they are consistent in their speech.

The training period will generally be longer and it may be necessary to train words individually to match the speaker's pronunciation. Speech recognition systems are generally able to cope with regional accents, although initial results may be disappointing and the training period will again be longer.

5.2 Literacy Skills

Given the frequent need to choose the desired word from a list of choices, speech recognition will be most useful for users with reasonably reliable word recognition skills. Discrete speech systems are easier to use for this than continuous ones since attention is focused on one word at a time. SR systems, by themselves, are not suitable for non-readers.

5.3 Cognitive Skills

The cognitive load involved in using speech recognition systems can be quite high. The person using it must not only think about what they want to say, but also how to say it; they must monitor whether the words they used have been recognized accurately, if not, they must decide on an appropriate strategy to correct them. That is not to mention the possibility of formatting text as it is being dictated. It is not surprising that people using speech recognition often find it tiring! People with conditions such as ME who find prolonged keyboard use tiring should bear in mind the likelihood that the use of speech recognition could be equally tiring - though in a different way. It would be a good strategy to switch between the two input methods. It is not possible to give some minimum level of cognitive ability below which SR will not be suitable So many other factors such as motivation and must also be taken into account.

5.4 Motivation

This is probably the most important factor for most people who try to use a speech recognition system. Initial results are often disappointing, particularly in comparison with the manufacturers claims. There will also be occasions when levels of recognition will seem to drop for no apparent reason (usually a change in the microphone setup, although this will not be obvious). Unless the user is well motivated to use the system, working to overcome the difficulties that arise, it is very easy to give up before the system has had a chance to adapt to the users voice. It is often the case that speech recognition systems are used most successfully by people with a desire to write, which may well have been frustrated by years of failure in the past.

5.5 Visual Skills

A wide range of information is presented visually from the screen: the text that has been entered; choices for an unrecognized word or phrase; information on how the program is running; even basic information as to whether or not the microphone has been switched on. Nonetheless, people with a visual impairment can successfully use speech recognition systems; the best results have been obtained in combination with, an appropriate screen reader. An example of this is the use of Naturally Speaking with Keystone Screen Speaker. There is an inevitable increase in the cognitive load, with the user having to dictate to the computer, while simultaneously monitoring a synthesized speech playback of what the system thinks he/she has just said.

5.6 Support

The level of support required will vary from user to user. An experienced computer user who has developed RSI, for example, should require much less support than a young child with complex disabilities. Nevertheless, even the former would benefit from advice from an experienced user as to whether initially poor recognition levels might be due to a microphone problem, or some other factor. A young child would need a lot of support, particularly in the training process, and it is arguable whether any but the most able child under 11 could manage a SR system, even with good support.

6. Problem Statement

6.1 EXISTING SYSTEM

The OS – Windows 7 contains an inbuilt "Speech Recognition ", but with following constraints:

6.1.1 LIMITATIONS

6.1.1.1. Navigational Path

To operate any application using "speech Recognition", a complete /full navigational path must be specified.

Strictly slang specific. Accepts only words pronounced in proper US/UK slang, which may be difficult for Asian countries.

The proposed system is really an efficient one which overcomes the constraints specified in existing system.

For e.g. to open a word document, the navigation path is as follows: Start -> Programs -> MS office -> Word document.

6.2 PROPOSED SYSTEM

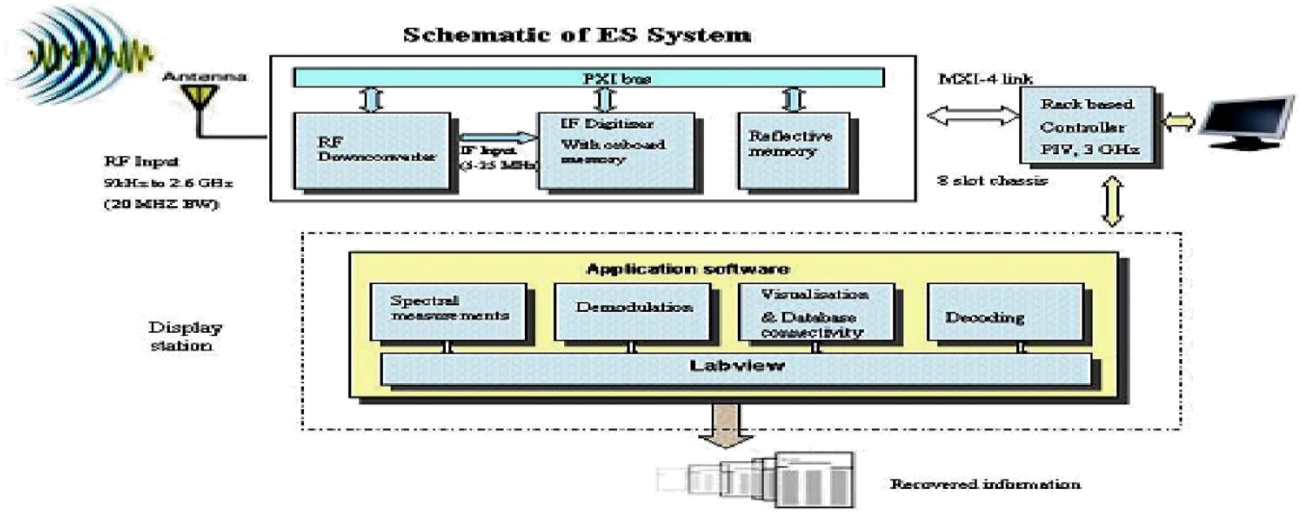


Fig.1. Architectural Diagram

6.2.1 ADVANTAGES

6.2.1.1 Navigational Path Constraint is overcome:

The proposed system contains a User Interface from where all operations can be done without following the navigational path.

6.2.1.2 US/UK English slang problem will be tried to overcome by implementing the system in Indian Languages.

6.2.1.3 The new proposed system will be developed as a separate software installer package, thereby allowing it to be installed on all windows machine.

7. Architecture-Block Diagram

USER uses MIC to say and execute a command. This MIC May be an independent MIC or HEADSET with MIC, also the MIC should be physically connected to

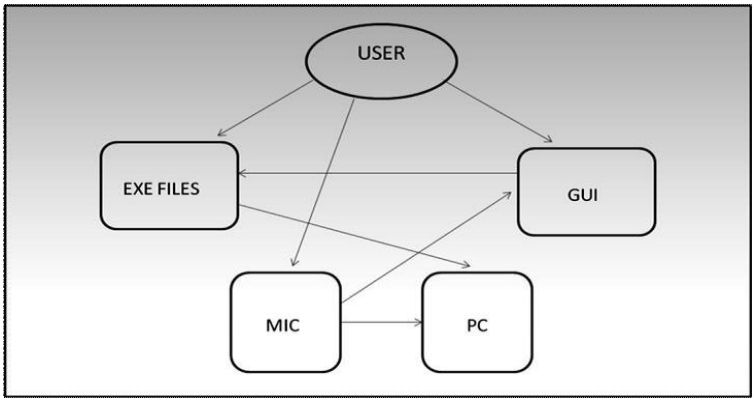


Fig.3. Usecase Diagram

7.1 DESCRIPTION OF ARCHITECTURE-BLOCK DIAGRAM

USER creates EXE FILES and GUI using C# DOT NET in "Visual studio 2005".

PC.MIC operates On GUI. GUI or FRONT END accepts input from USER through MIC, and invokes the EXE FILES. EXE FILES then controls the PC (i.e. certain system applications).

8. Class Diagram

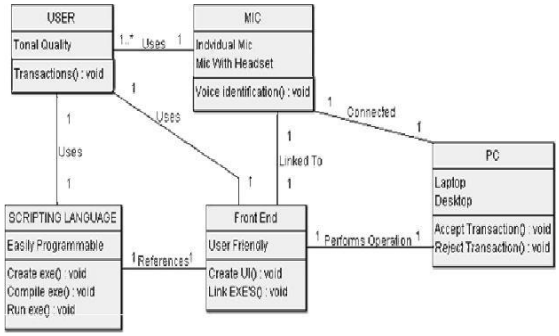


Fig.2.Class Diagram

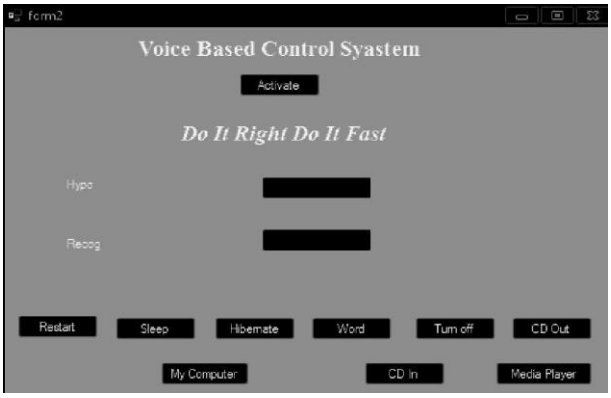
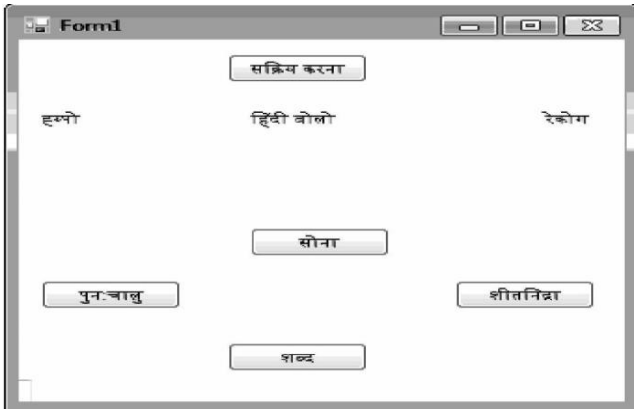


Fig.4. English GUI

9. Results & Screenshots

Fig.5.Hindi GUI



10. Conclusions

This research overcomes some common problems suffered by developers when bridging an application system to the interface of a recognizer. The proposed approach presents a more flexible and efficient interfacing. As a result, the proposed work enhances the interfacing of applications to recognizers by making it an easy, generic and flexible process. The major contributions of this study include:

1) Offers a simplistic and personalized way to interface applications with recognizers through the front-end, without the need of dealing with low-level issues such as system design and programming.

2) Allows modifications to a recognition interfacing environment of an application without requiring the access to source code of applications and re-compilation of it. Offers a generic and custom interface interfacing environment that allows the coexistence of multiple applications that hold different interfacing requirements

11. References

1. B. Balentine, D. Morgan, and W. Meisel, How to Build a Speech Recognition Application, Enterprise Integration Group, 1999.
2. VSpeech 1.0, Team BK02 product, http://vspeech. source forge.net.
3. R. W. Sebesta, Concepts of Programming Languages, 5th ed., Addison-Wesley Publishing Company, 2002.
4. Microsoft Speech SDK, Version 5.1 Documentation, Microsoft Corporation, 2001.
5. Shih-Jung Peng, Jan Karel Ruzicka and Deng-Jyi Chen,"A Generic and Visual Interfacing Framework for Bridging the Interface between Application Systems and Recognizers", Journal of Information Science and Engineering(JISE), Vol. 22, No.5, September 2006, pp.1077-1091.
6. J. K. Ruzicka, "The design and implementation of an interfacing framework for bridging speech recognizer to application system," Master Dissertation, Department of Computer Science and Information Engineering, National Chiao Tung University, Taiwan, 2005.