International Journal of Contemporary Research in Computer Science and Technology (IJCRCST) e-ISSN: 2395-5325
Volume2, Issue 3 (March '2016)

VOICE COMMAND EXECUTION WITH SPEECH RECOGNITION AND SYNTHESIZER (A VOICE INTERFACE)

Kalpana.A.V,
Assistant Professor,
Department of Computer Science and Engineering,
Velammal Institute of Technology,
Chennai,India.

Ram Kumar.C,
Student,
Department of Computer Science and Engineering,
Velammal Institute of Technology,
Chennai,India.

Avinash.R,
Student,
Department of Computer Science and Engineering,
Velammal Institute of Technology,
Chennai,India.

Venketaramanan.B,
Student,
Department of Computer Science and Engineering,
Velammal Institute of Technology,
Chennai,India.

Abstract: Speech technology is one of the fastest growing modern engineering technology with a wide scope for application in various arenas and disciplines of life. It has many potential benefits and is useful to people in many walks of life. Nearly 20% of people of the world are suffering from various disabilities; many of them are blind or unable to use their hands effectively. The speech recognition systems inthose particular cases provide a significant help to them, so that they can share information with people by operating computer through voice input.This project is designed and developed keeping that factor into mind, and a little effort is made to achieve this aim. Our project is capable to recognize the speech and convert the input audio into text; it also enables a user to perform operations such as "open and close applications and windows, select text, media controls, read text, system termination, social interaction"etc.,by providing voice input. It also helps the user to open differentsystem software such as opening MS-paint, notepad and calculator. At the initial level effort is made to provide help for basic operations as discussedabove, but the software can further be updated and enhanced in order to cover more operations.

I.INTRODUCTION

a) Objective:

- To understand the speech recognition and its fundamentals.
- Its working and applications in different areas
- Its implementation as a desktop Application
- Development for software that can mainly be used for:
- Speech Recognition
- Speech Generation
- Text Editing
- Tool for operating Machine through voice

b) Description:

Computers are known to execute accurately every command given to it by the user. There are various commands that can be inputted to the computer. The command could be in various formats. For example the command could be to print a document or to play audio/video files or to open a file or to paint a picture etc. These commands which are present in the user interface by default are implemented with the help of mouse pointer. The mouse pointer is used to select a variety of instructions listed which can perform the desired operation. The pointing of and selection by the mouse pointer consumes pretty much time although it is easier to use. Hence a need arises to curb the execution time and enhance efficiency of the system. To overcome this many new technologies have been proposed and implemented. One of the technology created is speech recognizer. It is a well-known fact that oral commands are obeyed and executed. Hence to make the computer recognize commands the speech recognizer was created. This feature is found in higher versions of windows operating system and in some high end mobile devices. This project has the speech recognizing and speech synthesizing capabilities and will be a good interface to be used through voice.

c) Existing System:

The above said speech recognizer is found in windows 7 and windows 8 as a secondary option. But it is not used upto its full capacity and capability. The application of this inbuilt speech recognizer is limited but it has many scopes. This operates with 3 steps. The user has to setup the recognizer. After the successful installation and setup of the speech recognizer the speech recognition engine is displayed on the desktop. To use the recognizer a command "show numbers" is to be given. Then many numbers at random will be shown by default. The user has to then select the number to perform the desired operation. This obviously increases the time of execution. This is the drawback of this system

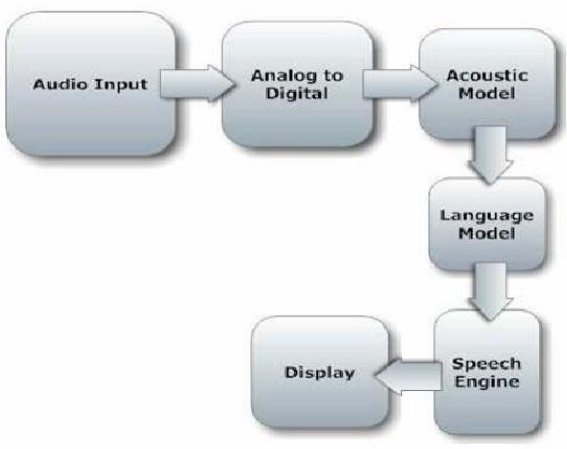
d) Proposed System

In the proposed system we have added many controls to the speech recognizer. The time of execution is reduced very much. All the commands are executed in a single step. Most of the basic computer tasks can be performed. This is found to be more efficient than the existing system.

II. SPEECH PROCESS:

- With the help of microphone audio is input to the system, the pc sound card produces the equivalent digital representation of received audio.
- The process of converting the analog signal into a digital form is known as digitization it involves the both sampling and quantization processes. Sampling is converting a continuous signal into discrete signal, while the process of approximating acontinuous range of values is known as quantization.
- An acoustic model is created by taking audio recordings of speech, and their text transcriptions, and using software to create statistical representations of the sounds that make up each word. It is used by a speech recognition engine to recognize speech. The software acoustic model breaks the words into the phonemes.
- Language modeling is used in many natural language processing applications such as speech recognition tries to capture the properties of a language and to predict the next word in the speech sequence .The software language model compares the phonemes to words in its built in dictionary.
- The speech engine will then process the modelled language and output is given.

III.METHOD:



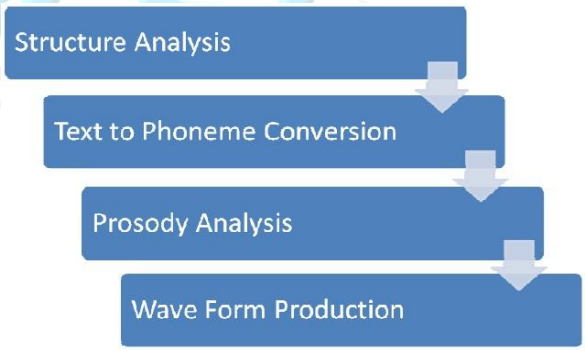
IV.METHODOLOGY

As an emerging technology, not all developers are familiar with speech recognition technology. While the basic functions of both speech synthesis and speech recognition takes only few minutes to understand (after all, most people learn to speak and listen by age two), there are subtle and powerful capabilities provided by computerized speech that developers will want to understand and utilize. Despite very substantial investment in speech technology research over the last 40years, speech synthesis and speech recognition technologies still have significant limitations. Most importantly, speech technology does not always meet the high expectations of users familiar with natural human-to-human speech communication. Understanding the limitations - as well as the strengths - is important for effective use of speech input and output in a user interface and for understanding some of the advanced features .An understanding of the capabilities and limitations of speech technology is also important for developers in making decisions about whether a particular application will benefit from the use of speech input and output.

Tools used:

1. Speech recognition engine
2. Microsoft Visual C# 2010
3. Notepad
4. Ivona reader

Synthesizer:



Speech Synthesizer: A speech synthesizer converts written text into spoken language. Speech synthesis is also referred to as

Text-to-speech: (TTS) conversion. The major steps in producing speech from text are as follows:

Structure analysis: Process the input text to determine where paragraphs, sentences and other structures start and end. For most languages, punctuation and formatting data are used in this stage.

Text pre-processing : Analyze the input text for special constructs of the language. In English, special treatment is required for abbreviations, acronyms, dates, times, numbers, currency amounts, email addresses and many other forms. Other languages need special processing for these forms and most languages have other specialized requirements. The remaining steps convert the spoken text to speech.

Text-to-phoneme conversion: Convert each word to*phonemes*. A phoneme is a basic unit of sound in a language. US English has around 45 phonemes including the consonant and vowel sounds. For example, "times" is spoken as four phonemes "t ay m s". Different languages have different sets of sounds (different phonemes). For example, Japanese has fewer phonemes including sounds not found in English, such as "ts" in "tsunami".

Prosody analysis: Process the sentence structure, words and phonemes to determine appropriate *prosody* For the sentence. Prosody includes many of the features of speech other than the sounds of the words being spoken. This includes the pitch (or melody), the timing (or rhythm), the pausing, the speaking rate, the emphasis on words and many other features. Correct prosody is important for making speech sound right and for correctly conveying the meaning of a sentence.

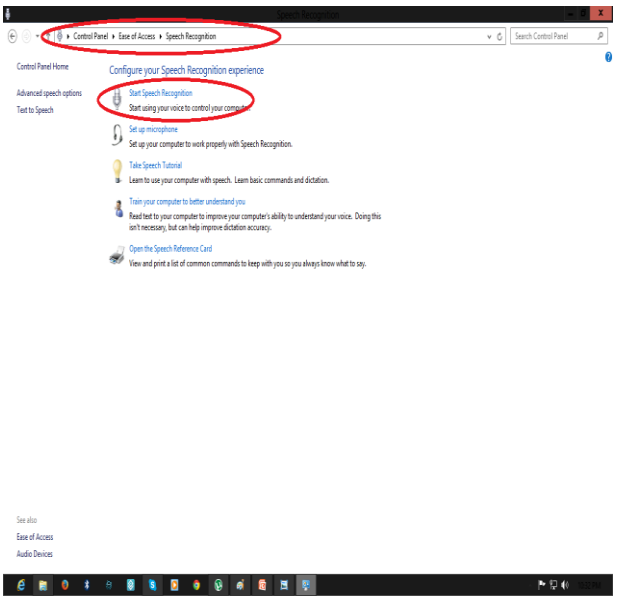
Waveform production: Finally, the phonemes and prosody information are used to produce the audio waveform for each sentence. There are many ways in which the speech can be produced from the phoneme and prosody information. Most current systems do it in one of two ways: *Concatenation* Of chunks of recorded human speech, or *Formant synthesis* Using signal processing techniques based on knowledge of how phonemes sound and how prosody affects those phonemes. The details of waveform generation are of no trivial use to users.

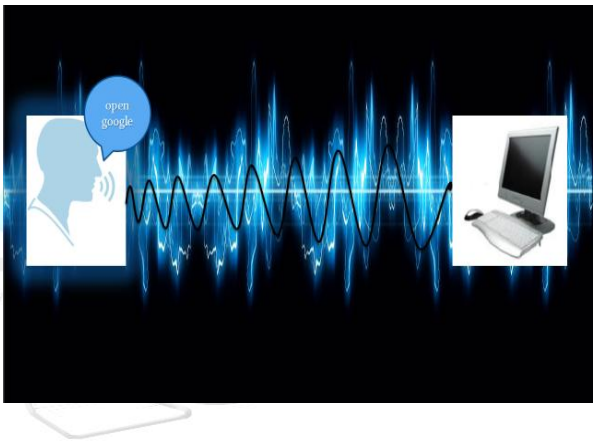
Speech Engine: The job of speech recognition engine is to convert the input audio into text toaccomplish this it uses all sorts of data, software algorithms and statistics. Its first operation is digitization as discussed earlier, that is to convert it into a suitable format for further processing. Once audio signal is in proper format it then searches the best match for it. It does this by considering the words it knows, once the signal is recognized it returns its corresponding text string.

V.WORKING:

This software is designed to recognize the speech and also has the capabilities for speaking and synthesizing means it can convert speech to text and text to speech.

The user is asked to provide voice command via the microphone. The microphone intakes the command and the analog signals are converted to digital ones in the internal circuit. These digitized signals are processed as acoustic model. The windows grammar verifies the command as a valid one in its default language. Then the speech recognition model comes into act.The speech recognizer application in windows 8 is connected through .NET in visual basics where the operational code is written in C# and Visual Basic invokes the application in frontend. Once the command is identified the application contemplates the command with the inbuilt code to execute the corresponding function. The program is essentially executed at run-time.

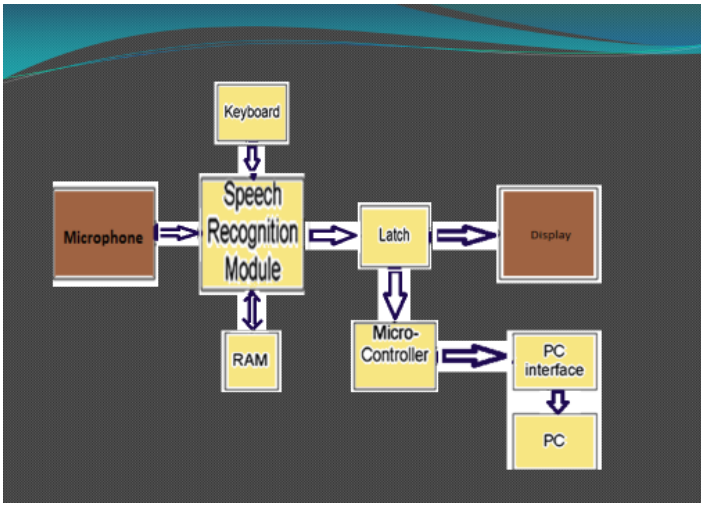




Through Voice Control, the computer uses both video and voice prompts to request input from the operator. The operator is allowed to enter data and to control the software flow by voice command or from the keyboard or mouse. The Voice Control system allows for dynamic specification of a grammar set, or legal set of commands. The use of a reduced grammar set greatly increases recognition accuracy. The computer voice enables the operator to focus his attention away from the computer screen, which is required for activities such as probing a circuit card and taking readings. When the operator takes readings, the computer, to insure reliable entry, echoes his voice entries. With electronic tuning, speech synthesis allows the operator to hear the resulting reading, enabling him to focus on the circuit card instead of constantly turning his head to see the computer screen. This projectenhances the capability and functionality of the Voice Control system.

The synthesize part of this software helps in verifying the various operations done by user such as read out the written text for user also informing that what type of actionsa user is

doing such as saving a document, opening a new file or opening a file previouslysaved on hard disk



Result of testing:

OPERATION	MODE OF EXECUTION	**RESULT**	INFERENCE
Open file/folder	a)keyboard/mouse b)voice input	Proper functioning of open,close operations Proper opening and closing of files and folders	While providing voice commands to open files after selecting them all files were opened/closed without errors.
Media player controls	a)mouse b)voice commands	Proper working of media player functions. Play, pause, slow,fast etc. worked without error.	Voice commands worked only for pause/play/stop. But could not switch between menu bar tabs. Not so accurate.
Text editing	a)mouse/keyboard b)voice commands	All functions work properly. Select,cut,copy,paste function accurately	Voice commands control editing functions.
Google search	a)voice commands	Can search any topic. can open new tab in the same browser. Can read webpages. Can go to next link. Can move to next page. Can open new webpage.	Browser functions work accurately without any hackles.
Reading selected text	voice	Voice commands are used to select and read the selected text without the use of mouse and keyboard.	Works good
Social interaction	voice	Interaction between computer and the user.	We provided predefined interaction commands. Answers to our queries properly if recognized the query properly.

VI.APPLICATIONS:

- A voice-controlled human-computer interface has been designed that enables severely handicapped individuals to operate a computer.
- **Hands-free computing** is any computer configuration where a user can interface without the use of their hands, an otherwise common

requirement of human interface devices such as the mouseand keyboard.

- Speech recognition systems can be trained to recognize specific commands and upon confirmation of correctness instructions can be given to systems without the use of hands.
- This may be useful while driving or to an inspector or engineer in a factory environment.
- Disabled persons may find hands-free computing important in their everyday lives. Just like visually impaired have found computers useful in their lives.

VII.CONSTRAINTS:

- Different voices cannot be recognized accurately.
- Requires in-depth coding.
- Needs to be customized on individual computers.
- Development of universal software is quite laborious.

VIII.FUTURE DEVELOPMENTS:

- Integration with other applications.
- Document typing.
- Enabling all commands present.
- Icon level implementation in all applications.

IX.CONCLUSION:

This Thesis of Project work of speech recognition started with a brief introduction of the technology and its applications in different sectors. The project part of the Report was based on software development for speech recognition. At the later stage we discussed different tools for bringing that idea into practical work. After thedevelopment of the software finally it was tested and results were discussed, few deficiencies factors were brought in front. After the testing work, advantages of the software were described and suggestions for further enhancement and improvement were discussed.

REFERENCES:

[1] Y. Yuan, "Relationship Between the Internet of Things and Consumer Electronics," IEEE Consumer Electronics Magazine, p.23, Apr. 2012.

[2] J. Decuir, "Introducing Bluetooth Smart: Part II: Applications and updates," IEEE Consumer Electronics Magazine, pp.25-29, Apr. 2014.

[3] J. Han, J.K. Yun, J.H. Jang and K.R. Park, "User-Friendly Home Automation Based on 3D Virtualb World," IEEE Trans. on Consumer Electron., pp.1843-1847, Aug. 2010.

[4] K. Balasubramanian and A. Cellatoglu, "Analysis of remote control techniques employed in home automation and security systems," IEEE Trans. on Consumer Electron., Vol.55, Issue 3, pp.1401-1407, Aug.2009.

[5] T. Kim, H. Lee and Y. Chung, "Advanced Universal Remote Controller for Home Automation and Security," IEEE Trans. on Consumer Electron., pp.2537-2542, Vol. 56, Issue 4, Nov. 2010.

[6] Y.R. Chuang, W.J. Yang S.J. Lin and T.L. Chiu, "Study and implementation of the smallest closed-area (SCA) mechanism for selforganization network architectures in smart home control systems," IEEE International Symposium on Consumer Electronics, pp.79-80, Jun. 2013.

[7] I. I. Papp, Z. M. Saric and N.D. Teslic, "Hands-free Voice Communication with TV," IEEE Trans. on Consumer Electron., pp.606- 614, Vol. 57, Issue 2, May 2011.

[8] J. S. Park, G. J. Jang, J. H. Kim and S. H. Kim, "Acoustic Interference Cancellation for a Voice-driven Interface in Smart TVs," IEEE Trans. on Consumer Electron., pp.244-249, Vol. 59, Issue 1, Feb. 2013.

[9] R. I. Damper, M. A. Tranchant and S. M. Lewis, "Speech versus keying in command and control: effect of concurrent tasking," International Journal of Human-Computer Studies, pp.337-348, Vol. 45, issue 3, Sep. 1996.

[10] K. M. Lee and J. Lai, "Speech Versus Touch: A Comparative Study of the Use of Speech and DTMF Keypad for Navigation," International Journal of Human-Computer Interaction, pp.343-360, Vol. 19, issue 3, Jan. 2005.

[11] Y. Yamazaki, Y. Fujita, and N. Komatsu, "CELP-based Speaker Verification: An Evaluation under Noisy Conditions," IEEE Internation conference on Control, Automation, Robotics and Vision, pp. 408-412, Dec. 2004.

[12] T.F. Quatieri, R.B. Dunn, D.A. Reynold, J.P. Campbell and E. Singer, "Speaker recognition using G.729 speech codec parameters," IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 2, pp. 1089-1092, Jun. 2000.

[13] H. Sakano, N. Mukawa and T. Nakamura, "Kernel Mutual Subspace Method and Its Application for Object Recognition," Electronics and Communications in Japan, Vol.E88, No.6, pp. 45-53, Jun. 2005.

[14] ITU-T, "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP)," ITU-T Recommendation G.729, 1996.

[15] K. Maeda and S. Watanabe, "A Pattern Matching Method with Local Structure," IEICE TRANS. on Information and Systems (Japanese Edition), vol.J68-D, no.3, pp.345-352, Mar. 1985.