Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICONECT-2015 Conference Proceedings

Intelligent Voice Operating System (IVOS)

Devang Sawant, Prateek Phadtare, Chirag Vyas
BE, Computer Engineering
KC College of Engineering and Management Studies
& Research
Thane

Abstract — **Microsoft and other OS designers, have tried their level best to incorporate features in their OS which can help differently-abled or handicap users to get the maximum out of their PC. However, these accessibility features have their own limitation, and they rarely allow the user to interact with applications other than OS itself. Example using media player or MS-Office is not easily possible. It is not possible to use all the features of the computer by people suffering from dexterity issues, motion impairment and blindness. So, for this people to use all the computer features to the fullest we will try our level best to design a super layer OS called intelligent voice operating system (IVOS). With the help of this system we will be able to operate, open, browse all computer applications by our voice commands. More applications can be added as and when needed and can be operated using voice.**

Keywords— Hidden markov model, Dynamic time warping, neural networks.

I. INTRODUCTION

Hardware and software that help people who are physically challenged, often called "accessibility options" when referring to enhancements for using the computer, the entire field of assistive technology is quite vast and even includes ramp and doorway construction in buildings to support wheelchairs. Enhancements for using the computer include alternative keyboard and mouse devices, replacing beeps with light signals for the deaf, screen magnifiers and text enlargers and systems that form tactile Braille letters from on-screen text. Environmental control units (ECU's), Automatic door openers, Dragon NaturallySpeaking voice recognition software for those with physical challenges.

Mac OS X includes a wide variety of features and assistive technologies known as Universal Access that include screen and cursor magnification, a full-featured screen reader, visual flash alerts, closed-captioning support, and much more. Mac OS X includes all the features your application needs to make it accessible to users with special needs.

Apple strongly encourages developers to support these APIs in all of their applications so they are compatible with features built into Mac OS X such as Voice-Over, as well as other third-party products. The X-code and Interface Builder tools, as well as the Cocoa frameworks, make it easy to add accessibility tags like roles and descriptions.

For example, Interface Builder has an Inspector that allows you to enter a description for any control in the user interface; that description will be synthesized into speech when Voice-Over is enabled.

1.1 How They Work:

An assistive application interacts with accessibility objects in application to allow people with disabilities to drive the user interface in non-traditional ways. For example, a Voice-Over user relies solely on the keyboard for control, and on Speech Synthesis for feedback. If your application can only be used with a mouse, it will be inaccessible to a user who relies on Voice-Over and other applications that use the Accessibility APIs in Mac OS X. Users with low vision can also set Mac OS X's built-in zoom, gray scale, and white-on-black display mode options to adapt the onscreen experience to their specific needs. Those who are deaf or hard of hearing can set audible alerts to automatically flash the screen instead.

Physically disabled users will rely on AppleScript and 'Automator' workflows to simplify complex tasks and can take advantage of keyboard and mouse preferences to make them easier to control and use. It is important that your application work as intended for those who rely on these assistive features of Mac OS X.

1.2 How IVOS Works:

IVOS (Intelligent Voice Operating System) is an intelligent agent that offers both Speech Recognition and Text-to-Speech capabilities, allowing you to run computer via voice commands. The menu and sub menu of any software, including total voice operation of MS Outlook can be controlled. You can use voice commands to open files, folders or websites and much more. In addition, the Text-to-Speech features allow you to convert your spoken text into typed letters, compatible with any software that offers a text input window (email messages, forms etc.). Additional features include dynamic voice control, dictation, transcription of recordings and more.

II. LITERATURE SURVEY

Voice based applications are taking over the market. Text will be replaced by voice in near future. Voice provides good authentication in comparison to text. In future all computer and mobile applications will be accessed by voice which will remove the need to use hands to deal with any handheld system.

2.1 Mitali Patil has written a paper on **The Design and Implementation of Voice Controlled Wireless Intelligent Home Automation System Based on ZigBee.** In this paper she has described that Intelligent Home Automation Systems are gaining importance in today's technology dependent world. Home Automation Systems provide a sense of security and comfort. Using Wireless technology like ZigBee the cost of wiring of Home Automation System can be reduced as well as a reliable and secure communication can be achieved. ZigBee is a low data rate wireless network standard with added features like low-cost, low power consumption and fast reaction. ZigBee is most suitable for small area networks like homes. This System also allows controlling of devices using Voice commands which reduce user interaction with system directly.

2.2 Ai-ping xiao has written a paper on **human machine interaction based on voice**. In this paper he has described that human machine system holds important position in every field of man and machine. It is also a key technology for a man to operate one machine. Voice based identification is one popular way to contact the user and device. It has been widely used in many areas with development of hardware technologies and improvement of mathematical algorithms.

2.3 Spion miah has written a paper on **To Design Voice Control Keyboard System using Speech Application Programming Interface**. In this paper With the passing of days men are more dependent on electronic devices. The Main objective of this project is to design and develop a voice Control Keyboard Systems, fully controlled by a computer, and display output on the display device with predefined time. So this project will work as a helping system for those person who has small knowledge about computer system even those person who are illiterate they can operate computer system. We can implement this developed system in other system for example voice control car system.

2.4 Parwinder pal singh has written a paper on **Speech Recognition as Emerging Revolutionary Technology**. In this paper he has described that Speech recognition is the translation of spoken words into text. It is also known as "automatic speech recognition", "ASR", "computer speech recognition", "speech to text", or just "STT". Speech Recognition is technology that can translate spoken words into text. Some SR systems use "training" where an individual speaker reads sections of text into the SR system. These systems analyze the person's specific voice and use it to fine tune the recognition of that person's speech, resulting in more accurate transcription.

2.5 G.Krishna has written a paper **on Voice based hardware controller**. In this paper he has described that The day is not far away, where, computer's peripherals performing tasks by taking commands from the most natural form of communication the human voice. Yes, this idea of developing a voice operable hardware device by using the speech recognition technology is to develop a basic application which demonstrates a hardware chip responding to the commands given by voice along with a GUI. Several years of research has already been done on the speech recognition technology and in our exploration we found that all the speech-recognizing engines were based on the HMM concepts although they may be written in different programming languages. HMM is the "Hidden Markov Model". This concept aims at building and manipulating Hidden Markov Model (HMM). HMM is primarily used for speech recognition applications. After exploring the options available for selecting a speech engine, which recognizes words, the best results were observed in a program written in Java. To show the hardware operation we have made use of the 8051 micro controller and embedded "C" code for its operation. So by using this as foundation we can build many applications like for ex: operating a printer through voice and any other output device connected to the computer. The advantage of our work in this paper is that our application is speaker-independent that can recognize continuous human speech regardless of the speaker and that can continually improve their vocabulary size and recognition accuracy.

2.6 V. Rudzionis has written a paper on **Voice-based Human-Machine Interaction Modeling for Automated Information Services**. The main aim of telecommunications is to bring people thousands miles apart, anytime, anywhere together to communicate as if they were having a face-to-face conversation in a ubiquitous tele-presence way. One key component necessary to reach this main aim is the technology enabling usual communication by voice. This means the use of automatic speech recognition. An IVR (Interactive Voice Response) based systems can be used to automate a wide range of services and data requests. These systems are used most often by the companies to provide the self-service abilities to customer. The system takes the input from the user and provides back the enterprise information in the form of recorded or synthesized voice, fax or even an email by connecting one or more online databases to the caller. Although there are several hundred million Internet-connected PCs in the world, this figure is dwarfed by the two billion fixed and mobile phones. The telephone is ubiquitous, increasingly mobile and could, in principle, provide a universal platform for accessing on- line services. To date efforts to harness this potential in the form of IVR systems have not proved especially popular with users. There's a wind of change blowing through the IVR world, impelled by advances in speech recognition technology and a transformation of the IVR programming environment.

2.7 G. Kokkinakis has written a paper on **A Speech-Based Human-Computer Interaction System for Automating Directory Assistance Services**. In this paper he described that the automation of Directory Assistance Services (DAS) through speech is one of the most difficult and demanding applications of human-computer interaction because it deals with very large vocabulary recognition issues. In this paper, we present a spoken dialogue system for automating DAS.1 Taking into account the major difficulties of this endeavor a stepwise approach was adopted. In

Special Issue - 2015

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
ICONECT-2015 Conference Proceedings

particular, two prototypes D1.1 (basic approach) and D1.2 (improved version) were developed successively. The results of D1.1 evaluation were used to refine D1.1 and gradually led to D1.2 that was also improve using a feedback approach. Furthermore, the system was extended and optimized so that it can be utilized in real-world conditions. We describe the general architecture and the three stages of the system's development in detail. Evaluation results concerning both the speech recognizer's accuracy and the overall system's performance are provided for all prototypes. Finally, we focus on techniques that handle large vocabulary recognition issues. The use of Directed Acyclic Word Graphs (DAWGs) and context-dependent phonological rules result edin search space reduction and there for ein fast erresponse and also in improve accuracy.

2.8 H. James Landay has written a paper on **Longitudinal Study of People Learning to Use Continuous Voice-Based Cursor Control**. In this paper he has described that We conducted a 2.5 week longitudinal study with five motor impaired (MI) and four non-impaired (NMI) participants, in which they learned to use the Vocal Joystick, a voice-based user interface control system. We found that the participants were able to learn the mappings between the vowel sounds and directions used by the Vocal Joystick, and showed marked improvement in their target acquisition performance. At the end of the ten session period, the NMI group reached the same level of performance as the previously measured "expert" Vocal Joystick performance, and the MI group was able to reach 70% of that level. Two of the MI participants were also able to approach the performance of their preferred device, a touchpad. We report on a number of issues that can inform the development of further enhancements in the realm of voice-driven computer control.

2.9 I. Gerad Chollet has written a paper on **Multimodal Human Machine Interactions in Virtual and Augmented Reality.** In this paper he has described that Virtual worlds are developing rapidly over the Internet. They are visited by avatars and staffed with Embodied Conversational Agents (ECAs). An avatar is a representation of a physical person. Each person controls one or several avatars and usually receives feedback from the virtual world on an audio-visual display. Ideally, all senses should be used to feel fully embedded in a virtual world. Sound, vision and sometimes touch are the available modalities. This paper reviews the technological developments which enable audio-visual interactions in virtual and augmented reality worlds. Emphasis is placed on speech and gesture interfaces, including talking face analysis and synthesis.

III. METHODOLOGY

Voice recognition consists of two main processes: acquiring speech signals, and processing the signals with computer algorithms to remove background noise and detect the speech accurately. Acquired signals can be used to manipulate different actions, such as the rejection of background as well as white noise, to follow the command of the user, or accurately move the object such as wheelchair upon user's wish.

In voice recognition, numbers of DSP algorithms are used to process the speech signal. Often preloaded libraries

intelligently predict the future words and complete the word/sentence based upon user's initial words. Speech recognition is generally implemented using Voice Activity Detection (VAD) for start and end detection, as well as zero crossing method and 4th order cumulants to determine the presence of speech. In order to achieve a quality speech signal, the bit rate and the sampling frequency of input signal should not be exceedingly high.

In case of speech detection for dysarthrias patients, the overall algorithm and process become more complex due to difference in energy and frequency of tone. Some problems associated with the speech of the dysarthrias due to neuromuscular deficiency are "velpharygeal noise, irregular articulation breakdown and mispronunciation of the fricative /v/ as the nasal /m/ ."

IV. BLUDING BLOCKS FOR IMPLEMENTATION

Both a strong software and hardware are necessary to implement a speech detection system. The first element required for speech detection is DSP algorithm as well ADC/DAC. The microphone is connected at the input of the system where speech signal is detected. The input signal is send to the signal processing CMOS chip, where the DSP algorithm is performed on the speech signal, and finally the signal is outputted to the speaker which is connected to PCB board using USB. Since Matlab is slower compared to C/C++, DSP coding is mainly done in C++ programming for processing signals. Overall CMOS chip would include modules for acoustics, dictionaries, along with recognition decoder and ADC at the input end. Once the speech signal is detected, software will process the signal to accurately detect the speech and result would be outputted through D/A convertor. The load end of the system should be low resistance approximately 75-800 Ω in order to reduce the overall system power consumption, which would be helpful in powerless CPU system, or where the utilization of speech recognition systems has been limited by software . Or in other words, the input impedance should be 5-10 times higher compared to output impedance of the system.

On the hardware side, the main input element is the microphone. In order for microphone to supply a good speech signal to the ADC on the chip, it should meet important specification. Microphones respond to 20 Hz to 20 KHz frequencies better compare to higher frequencies. Sensitivity of the microphone should not exceed +/- 3dB. In addition, the voltage produced in responses to an acoustic stimulus should be in hundreds of mV/Pa in a microphone. For example, a sensitivity of 70 mV/Pa means the microphone produces an output of 70 mV when presented with an input of 1 Pascal (94 dB SPL).

4.1 Requirements Analysis

Following are the technology we will be using

1) VB6 (windows software development)

2) Windows XP/7 or 8.

3) The Speech API: has been an integral component of all Microsoft Windows versions since Windows 98. 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. Code written for SAPI 5.3 (Vista) will run on SAPI 5.4 (Windows 7) without recompiling.

4) Visual Studio 6: We will be using Visual Basic 6; which a part of Visual Studio 6 is. Although language like C and Java can be used we have decided to use Microsoft technology. Also VB provides more appealing GUI as compared to any other language. GUI for application like Photoshop is written in VB, and its mathematical computation in C.

V. ADVANTAGES AND FEATURES

We intend to provide following features:

- Internet browser
- MS-Office/ Excel / Word / PowerPoint enabled
- Windows Media Player
- Microsoft paint
- Customized calculator

Users will be able to add any application which they want to make voice enabled. Custom commands for that application (Save-As, Open, Insert) etc. can be created and added by the user. Any internet site can be added as an when required

VI. DISADVANTAGES

User may not be able to write a mail using free speech, however on-screen keyboard will help him with this kind off scenario

VII. FUTURE SCOPE

Human computer interaction in artificial intelligence is a promising field, IVOS can largely affect how users use their computer with minimum use of keyboards and mouse.

REFERENCES

[1] Davies, K.H., Biddulph, R. and Balashek, S. (1952) Automatic Speech Recognition of Spoken Digits, J. Acoust. Soc. Am. 24(6) pp.637 - 642

[2] Speech.kth.se

[3] Eurofighter Direct Voice Input

[4] "The History of Automatic Speech Recognition Evaluations at NIST". National Institute of Standards and Technology. May, 2009. http://www.itl.nist.gov/iad/mig/publications/ASRhistory/index.html. Retrieved May, 2010.

[5] Goel, V.; Byrne, W. J. (2000). "Minimum Bayes-risk automatic speech recognition". Computer Speech & Language 14 (2): 115–135. doi:10.1006/csla.2000.0138. http://www.clsp.jhu.edu/people/vgoel/publications/CSAL.ps. Retrieved 2011-03-28.

[6] Mohri, M. (2002). "Edit-Distance of Weighted Automata: General Definitions and Algorithms". pp. 957–982. doi:10.1142/S0129054103002114. http://www.cs.nyu.edu/~mohri/pub/edit.pdf. Retrieved 2011-03-28.

[7] "Speech recognition for disabled people". http://www.businessweek.com/1998/08/b3566022.htm.

[8] John Pierce (1969). "Whither Speech Recognition". Journal of the Acoustical Society of America.

[9] Janet M. Baker, Li Deng, James Glass, Sanjeev Khudanpur, Chin-Hui Lee, Nelson Morgan, Douglas O'Shaughnessy (MAY, 2009). "Research Developments and Directions in Speech Recognition and Understanding, Part 1". IEEE SIGNAL PROCESSING MAGAZINE. http://research.microsoft.com/pubs/80528/SPM-MINDS-I.pdf. Retrieved May, 2010.

[10] S. Suk, S. Chung, and H. Kojimma, "Voice/Non-Vocie Classification Using Reliable Fundamental Frequency Estimator for Voice Activated

[11] A. Little, and L. Reznik, "Speech Detection Method Analysis and Intelligent Structure Development," In Proc. Australian New Zealand Conference on Intelligent Information System '96, 1996, pp. 2. [Accessed Sept. 3, 2008]Ekpe Okorafor, Mensah Kwabena Patrick Real-time Streaming Analysis for Hadoop and Flume, 2012