

SPEECUR: INTELLIGENT PC CONTROLLER FOR HAND DISABLED PEOPLE USING NLP AND IMAGE PROCESSING

E.J.A.P.C. Chathumali¹, J.M.U.B. Jayasekera², D.C.Pinnawala³, S.M.T.K. Samaraweera⁴, A Gamage⁵

Faculty of Computing
Department of Information Technology
Sri Lanka Institute of Information Technology

¹chathujay92@gmail.com, ²ubjayasekera@gmail.com, ³cdevindu@gmail.com,
⁴kemindasamaraweera@gmail.com, ⁵anjalie.g@sliit.lk

ABSTRACT

Today computers play a major role in human lives. Even though there are very sophisticated interfaces, differently abled people find it challenging to interact with the computer. There are some applications for disabled people, but people who have disabilities in hands do not have a proper application to interact with new technologies. The aim of this project is to develop software that act as an intelligent controller to facilitate hand disabled people when interacting with a computer. Proposed intelligent solution is based on speech recognition, image processing and human computer interaction. This application is capable of moving mouse cursor with face detections, activities based on voice commands, provide user authentication by voice recognitions and give suggestions using facial emotions. The solution will be using various high end techniques in Natural Language Processing, Machine Learning and Image Processing in order to improve the computer interaction. The proposed solution will be a great solution for the hand disabled people to interact with the computers like a normal user does.

Key Words : Natural Language Processing, Facial point detection, Voice based Authentication

1. INTRODUCTION

Few years before computer technologies reached to an impressive level. With the advance of computer technology, the society has changed drastically. In the current situation everything is linked with computers in some way. As a result the computer has become a major contribution for the society to become a global village. Key board and mouse help people to interact with the computer mainly. Even though normal people can access those devices easily, it is difficult for hand disable people to use those devices and interact with the computer. Therefore in order to increase the computer literacy furthermore, it is vital to come up with a good solution for this situation. The aim of this project is, to design Intelligent PC Controller named "SpeeCur", to help the hand disabled people to interact with computers by Human Computer Interaction (HCI). The user interface (software) represents the point of communication between a computer system and a person (user). Among many researches done

by different people over the last few years, voice/speech recognition and face detection technique are the most suitable solutions for hand disabled people to interact with the computer.

Speech/voice and face are major natural methods of information exchange between human beings. As human beings, we speak and listen to each other in human-human interface. We communicate with each other by using facial expressions. For centuries people have tried to develop machines that can understand and produce speech as humans do so naturally. Voice/speech recognition is a field of computer science that deals with designing computer systems that recognize spoken words. It is a technology that allows a computer to identify the words that a person speaks into a microphone or telephone. Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. [1] Common speech recognition systems these days can recognize thousands of words. Speech

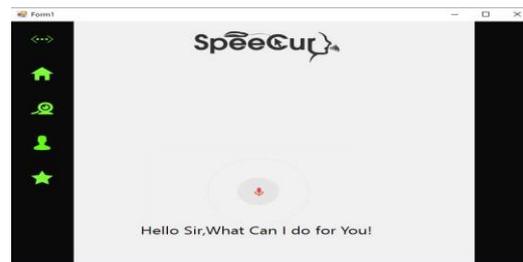
recognition is considered as an input as well as an output during the Human Computer Interaction design. Automatic speech recognition (ASR) is one of the fastest developing fields in the framework of speech science and engineering. Automatic speech recognition has an interesting and useful implementation in expert systems, a technology whereby computers can act as a substitute for a human expert. [2] An intelligent computer that acts, responds or thinks like a human being can be equipped with an automatic speech recognition module that enables it to process spoken information. Speech recognition system is capable of converting human speech to text or commands for the computer. Making and learning user preferences, detecting and processing audio signals plays great deal in this system. The recognized words can be the final results, as for applications such as commands & control, data entry, document preparation and so forth. They can also serve as the input to further linguistic processing in order to achieve speech understanding.

There are two main reasons for getting such applications popular, first is the wide range of commercial and social service, second one is directly interacting with the disable people(hand) and help to do their day today activities in a normal way. Most of the time disable people are hidden from the society and in post-perdition time many soldiers got retired before completing their time of service. The new system will combine many technologies such as face recognition and curser movement, voice recognition and voice command and facial expression and syncing with social media. Head and face detection is the first step to be taken for facial expression recognition and head gesture recognition. Real time face detection has evolved in the recent years with many methods, and algorithms been proposed for detection of face. The first challenge is to correctly detect the face with a webcam. The second challenge is to correctly detect speech. Many research projects and tools are invented to use voice recognition but after researching and studying about those systems several limitations were also identified.

2. LITERATURE REVIEW

Throughout the phase of literature survey several applications were found which use voice recognition techniques and face detection techniques to interact with the computer. Still there are lots of researches carried out on voice

recognition systems for computer interaction. *Kathana* is one such solution basically for recognizing and interpreting voice. [3] It converts an acoustic signal which represents human speech done in Sinhala language



captured by a microphone, to a set of words. Emphasis is that this acoustic wave represents a human speech done in Sinhala language. The recognized words which are the results can be used for applications as commands, data entries or could be served as the input to further linguistic processing in order to achieve tasks done in Sinhala language and produce an output that can be used for other applications extensively.

Siri is another such application developed by Apple which works as a personal assistant and knowledge navigator. The application uses a natural language user interface to answer questions, make recommendations, and perform actions by delegating requests to a set of Web services. The software, both in its original version and as an IOS application, adapts to the user's individual language usage and individual searches with continuing use, and returns results that are individualized. Siri has impressive human-like voice interaction. [4] Siri's primary technical areas focus on a Conversational Interface, Personal Context Awareness, and Service Delegation. Siri also contains numerous pre-programmed responses to conversational and amusing questions. These are designed to provide an entertainment factor and give Siri human-like qualities. [4]

E-Speaking is another comprehensive voice and speech recognition program to use voice for command & control of computer and dictation. [5] It is a shareware. Computer can be operated using minimum number of keyboard inputs and mouse clicks. Voice commands are used to open Web sites, documents, or programs. Also e-speaking perform navigation and editing functions, dictate letters, memos, and reading email messages. Voice Dictation is integrated into the e-Speaking application including 26 different Dictation Voice Commands. [5] Voice

commands and recognition application were developed utilizing Microsoft's Speech Application Program Interface (SAPI) and Microsoft's .NET Framework. SAPI and .NET are both required to use this software. *Tazti* is a speech recognition software package developed and sold by Voice Tech Group, Inc. for Windows personal computers. [6] Software utilizes a minimal user interface. User can set up *Tazti* just as they please. User can set up, edit and delete commands as they please. The major drawback of the software is, that it won't let user dictate text to a document. This software has four primary areas of functionalities, playing PC games, control PC based applications and programs by voice, open files, folders and webpages by creating custom speech commands and associating them to file, folder or web URL, run programs that include command line parameters. *Vocalise* is a Speaker recognition software developed by Oxford University Research Group.[7] This system of automatic speaker recognition is based on both phonetic and spectral parameters and is able to compare the audio file of the target speaker with a set of audio files that belong to the suspicious speakers and creating a sorted list of files that are closest to the voice of the target speaker. It also provides a list of similarity score that quantifies the closeness of the match. *Vocalise* Offers the unique capability of being able to perform comparisons using either features traditionally used in forensic phonetic acoustic analysis like pitch, formants, etc. as well as spectral features such as Mel Frequency Cepstral Coefficients (MFCCs). It is designed to have a straightforward and intuitive user interface, and by design does not implement any limitation on the duration of the recordings required or of signal quality and this is left to the user to decide. *Vocalise* applies techniques similar to those used in automatic speaker recognition to phonetic features. These include the use of a universal background model, consideration of formant dynamics and long term formant distributions, as well as GMM-modeling, feature scaling and normalizations.

Therefore according to the literature survey the proposed system provides more benefits than the existing systems. Because when considering about the features of the existing system, those are only sub functions of the proposed system. Researches can assure that the proposed systems will gain more for as a business process.

3. MEHODALOGY

The main objective of this project is to deliver a voice and facial cursor movement based solution for the disabled persons to use a computer. The main focus is to improve the issues that are not properly addressed by existing solutions. The outcome is an application where disabled users have the freedom to do anything on computer as a normal user.

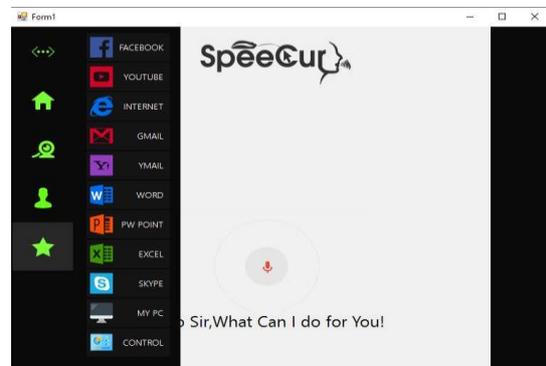


Figure 1 : Main Interface

The above Figure 1 provides the implementation results and solutions provided for the identified research problem. The main user interface of the basic implementation of the "SpeeCur" system is shown above.

3.1 Modules

1) Face Authentication.

The following methods were used for the Face Authentication Module

- Eigen Face Recognition
- Fisher Faces Recognition
- Local Binary Pattern Histograms Recogniton

2) Speaker Recognition.

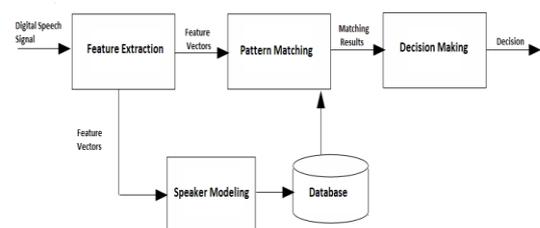


Figure 2: Speaker Recognition

- Speaker Identification - Indicates a speaker from the collection of well-known speakers, using his voice tag
- Speaker Verification - process of accepting or rejecting the identity claim of a speaker.
- Feature Extraction - helps to estimate variables by observing a speech signal. Main goal is to transform the data contained in the speech signal in to a variable collection that preserves these information and later can be used for comparison.
- Pattern matching - module compares or matches the extracted features with speaker models such as hidden Markov model, Gaussian Mixture Model, dynamic time wrapping model and vector quantization model. In this system it will consider Gaussian Mixture Model GMM.
- Decision model - finds similarity score between the given test sample and the claimed speaker to recognize the speaker. The acoustic features contain the characteristic information of the speech signal and are useful for recognizing the speaker.

3) Face Detection and Cursor Movements

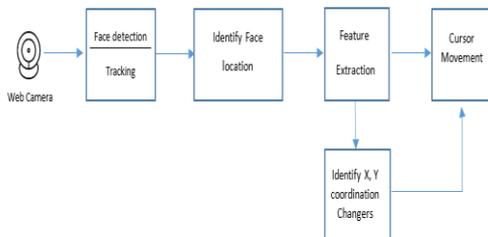


Figure 3: Speaker Recognition

4) Voice Recognition and Command Control.

With the speech component it can access to the speech recognition and speech synthesis which used to create the speech recognition engine and speech synthesis engine. Speech recognition engine is the main component which used to create the Grammar, Transferring the voice to the default audio device, identify the recognition mode and to create the recognition events of the

SpeeCur application. Speech Synthesis is used to create the synthesizer which can read the commands to the user in an intelligent way.

5) Emotion detection and Sugesstion Handling

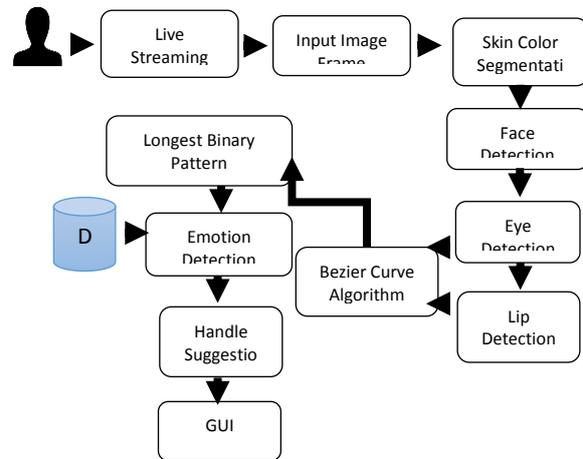


Figure 4: Process of Emotion Detection

4. DISCUSSION

In this section we sought to derive conclusions by recognizing the value of developing the SpeeCur system with comparing the percentage distribution of disabilities by type of disability and comparing the existing technologies with the SpeeCur. From the phase of literature survey several applications were found which use voice recognition techniques and face detection techniques to interact with the computer. Still there are lots of researches carried out on voice recognition systems for computer interaction. There are several existing technologies developed with the voice recognition and face recognition. But SpeeCur provides more efficient and robust features that the existing systems.

5. CONCLUSION

With the fast development of Technology, the computers deeply relate to the people in their day to day activities. So the computer users and computer related accessories are increasing day by day. Even though there are huge amount of computer users, differently abled people find it challenging to interact with the computer. There are some applications for disabled people, but people who have disabilities in hands do not have a proper application to interact with new developing technologies. But when considering

the proposed solution of ours it is a well-balanced application included with voice recognition and face recognition cursor movements to interact with the computer.

And when considering existing applications, they are lot of systems developed for the blind people. Not a particular solution for the people with hands disabilities. But the proposed system has a unique feature that having voice and face recognition both in a single system. User has given the facility to interact with the computer using voice and commands and move cursor using the facial movements. Although the SpeeCur has the unique features such as facial emotion recognition to identify the users working mood and facial user authentication to give an efficient user recognition.

So finally the outcome of our proposed system carries an efficient and robust application with voice recognition concepts and gesture recognition concepts which can be used by the people who are unable to use the hands to interact with computer. Ease of access and more user friendliness based desktop application is developed addressing the major issues faced by the people who are unable to use computer by their own hands. "SpeeCur" the voice and cursor movement based system sensitively concerns about all problem that faced by the targeted users and it will be a revolutionarily good solution for them within local context as well as globally.

REFERENCES

- [1] Zue, V., Cole, R., "Survey of the State of the Art in Human Language Technology" [Online] Available: www.dfki.de/~hansu/HLT-Survey.pdf. [Accessed: 05th January 2016]
- [2] Muhirwe Jackson, "Automatic Speech Recognition: Human Computer Interface for Kinyarwanda Language" [Online] Available: http://www.fon.hum.uva.nl/IFA-publications/Others/Courses/MuhirweJackson_final_report-1.pdf [Accessed: 12th February 2016]
- [3] Maleen Cooray, Tharindu Munasinghe, Chamal Rajapakse, Randika Yatawatte, "Kathana Sinhala Speech Recognition System" [Online] Available: <http://kathana.googlecode.com/files/Final%20Report%20-%20introductoryChapters.pdf> [Accessed: 23th January 2016]
- [4] Eroca Sadun, Steve Sande "Talking to Siri" [Online] Available: <http://ptgmedia.pearsoncmg.com/images/9780789750693/samplepages/0789750694.pdf> [Accessed: 02nd February 2016]
- [5] E-Speaking, "Control your PC Computer using Free Voice and Speech Software" [Online] Available: http://www.speaking.com/control_pc_computer_free_voice_speech_software.html [Accessed: 03th February 2016]
- [6] Tazti "tazti Speech Recognition by Voice Tech Group" [Online] Available: <http://www.tazti.com/products.html> [Accessed: 03th February 2016]
- [7] Speaker recognition with Phonetic and Automatic Features using VOCALISE Software. Anil Alexander, Oscar Forth, Marianne Jessen and Michael Jessen Oxford Wave Research Ltd [Online] Available: <http://www.oxfordwaveresearch.com/j2/products/vocalise?flatlia=1> [Accessed: 23th January 2016]
- [8] "Camera Mouse Including "Ctrl-Alt-Del"" http://www.ijcsmc.com/International_Journal_and_Adv._Comp_Sci_and_App, Vol. 4, No.3, pp.183-190, [Accessed 03rd March 2016].
- [9] "Graphic and Media Lab" [online] Available : http://www.bogotobogo.com/python/OpenCV_Python/python_opencv3_Image_Object_Detection_Face_Detection_Haar_Cascade_Cassifiers.php. [Accessed 03rd March 2016]
- [10] A. Mehrabian, "Communication without words", psychology today, vol. 2, no. 4, pp. 53-56, 1968 [Online] Available: <http://www.ijese.org/attachments/File/v1i7/G0325051713.pdf>, [Accessed 03th February 2016]