I J C T A, 9(15), 2016, pp. 7027-7033

Gesture Conversion to Speech: A Communication Module for the Speech Impaired

Khushboo Agarwal, Aditya, Anshul Nalin, Jaspreet Singh Riar and Sabitha Gauni

ABSTRACT

About 8% of India's population is multiply disabled of which major sections constitute of women due to lower literacy rate, social factors and lower grade health facilities. In this paper we review a few existing models to aid people with speech disability. Further we explore a novel approach based on gesture recognition to mitigate few shortcomings described above. Finally we give a solution which will address most of the problems enumerated in this paper to help the speech disabled people living in developing nations to uplift their daily living and also become economically independent which would in turn help the overall development of the nation.

Keywords: accessibility, speech impairment, gesture recognition

1. INTRODUCTION

Disability is seen as a part of the human condition where a person is deficient of one or more of his senses or efficient motor functions. Despite the rise in quality of healthcare, research in medical field to fight cancer, diabetes, etc. not much has been done to improve the conditions and environment for people with disabilities. Moreover there is a lack in services and research for people who are suffering from speech disabilities. According to the 2011 census of India a total of 26 million people suffer from some form of disability [1].

World Health Organization has estimated that over 1 billion people (or approximately 15 percent of the world's population) are living with some form of disability [2]. This number will increase due to ageing populations and the higher risk of disability in older people as well as the global increase in chronic health conditions. All over the world, people with disabilities have poorer health outcomes, lower education opportunities, less economic independence and higher rates of poverty than people without disabilities [2].

Out of the above-mentioned 15 percent, 8 percent live in developing nations [3]. Figure 1, shows the proportions of various disabilities among the disabled population. It also gives the shocking figure of 7.5% of disabled population to be speech disabled in India. The rates of disability are higher among the group with lower education than the group with higher education. In India, women constitute a higher vulnerability to disability than men. They are recognized to be multiply-disadvantaged, mainly due to the exclusion from the society on the basis of their gender and their disability [3].

With people becoming more conscious and realizing the need to uplift the quality of everyday living for disabled, there have been many devices to assist them. Many of the discussed devices in Section II have been modeled to assist a particular impairment. At the industrial level, the main reason for the failure of assistive devices is seen due to the high investment at the initial level and the market to be very limited. Our proposed design in further sections aims to solve this problem especially at the developing nation market's level.

* SRM University, Kattankulathur, Chennai, India, *Email: khushi.dps@gmail.com, riarjaspreet77@gmail.com, sabitha.g@ktr.srmuniv.ac.in*

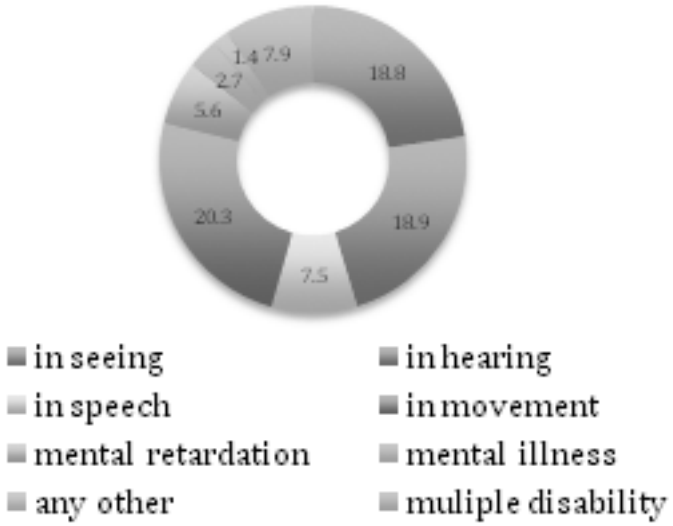

Figure 1: Proportion of Disabled Population by Type of Disability. India: 2011

The people who are usually mute (speech- disabled) are sometimes the luckiest of the lot or not. Lucky because their disability is not visible to everyone, but many a times this can become a bane too. The people, who are mute, usually have problems with body parts, which are used for verbal communication like esophagus, larynx, vocal cords, etc. Sometimes the speech of the person is impaired due intellectual disability or a side effect of a disease or treatment. The difficult situations faced by a mute individual are usually related to lack of expression, communicating emotions, the disability to call alarm, etc. Many devices have been invented to accommodate the speechlessness of people.

In Section II, this paper describes the previous work conducted by various researchers. Further Section III will focus on the project overview delineated in this paper in subsequent sections. Section IV and V will focus on the hardware and the software development and deployment of this project. Further developments and conclusions are under Section VI.

2. PREVIOUS WORK

Augmentative or alternate communication (AAC) is a term, which is usually related to indicate technologies that are used as an alternate for people who cannot speak or express themselves properly. There are two kinds of aids available, one is which doesn't use any tools and the other is the one, which utilizes tools (external means) to provide an alternate solution to people who cannot speak to enable them to express themselves. The unaided AAC uses techniques like sign language, gestures, visualizations and facial expressions whereas aided AAC utilizes electronic, tactile or visual means to transmit or receive messages [4].

As described above in [5, 6, 7, 8, and 9] sign language converters and interpreters can also be used for augmentative communication for people who cannot speak. Usually people who are speechless use sign language to communicate with other people. The problem which occurs in this situation is that either they know sign language or not, which has been overcome by a lot of sign converters described above. Methods in this era are required for communication between people who know sign language and those who do not.

In [10], an innovative way has been implemented using a personal computer where text is converted to sign language for those who do not know how to sign. A database table has been created and saved, which is used for correlation between the input texts and to output the signed language using a 3D sign hand. This way of communication helps to communicate between a speechless person and a deaf person also, as well as a person who knows sign language. A speech output depends on a lot of characteristics like the speed of speech, the quality of the sound, the accent, the language, etc. In [11], the researchers describe various

methods needed for speech visualization for the output like a database, filtering of sound, corresponding an input sound for the one in the database.

A person who cannot speak usually uses speech synthesizers in order to communicate with other people. Such a speech synthesizer has been described in [12], which enumerates speech quality, vocabulary, speech speed and cost as factors, which should be considered before putting a speech synthesizer in commercial use. It uses a pictorial language, which uses pictures as an input and the speech is given as an output. Some preset sentences exist such as, “I like”, “I want”, etc. so instead of lettered keys, apple, tea, leave, trees etc. are pressed to complete the sentences, which is given as an output. It also gives an access to various icons and different combinations so that it is easier for the user to form sentences. If the person is not able to move the cursor or pointer on his own, then a pointer control is mounted on a body part to be controlled by the communicator to point at icons, which can be used to complete the sentences, which are then spoken out using a speech synthesizer. This method is useful because the writing speed is overcome here even if the user lacks the manual dexterity.

To overcome the exclusion in social environment for the mute people and children, a game project is advocated to bring normalcy in their lives and improve upon the quality of lives they lead. This game is based on image processing where Kinect (a 3D- stereo camera) is used to track the players and gestures are used to relate to the text to be played [13]. This game does not require any voice input or hearing output, it is mainly based on the controlling the game environment using basic gestures, which are already preset for the game. This game would encourage inclusions into the society, uplift game spirit and boost up the confidence in a person.

Kinect has also been used for gesture recognition in [14]. It is described here to translate the sign language into the text language, which is available in the already loaded dictionary. Initially the filtering of image sequences is performed and soon after classification of gestures is done on the basis of already set classifiers and learning algorithms. Also new gestures or customized gestures can be individually stored by the user/communicator manually.

Some researchers have also proposed speech-generating devices under AAC rather than manual signing as an easier and more comfortable way of communication. Speech generating devices are of two types; the first type includes the output speech to be a sound of alarm or a particular recorded message, while the other type refers to a much more complicated speech synthesis and processing output. These devices also help in the case of developmental disorders. Speech generating devices are faster to use and also more interactive as the keys could be either color coordinated or various icons are placed on the keys to provide a better visual to the user [15].

3. PROJECT OVERVIEW

In this section, we illustrate the features and the projected working model of the project. In our system, we have tried to take up two approaches to solve the communication problem as described above among the speech challenged population of the world. Also any reference to sign language is to the American Sign Language merely due to the ease of acceptability; but final implementation of this project can utilize any other sign language convention easily.

3.1. Approach 1: Using flex sensors

In this approach, the user wears a special glove which is fitted with flex sensors. These sensors in coalition with software manipulation with the microcontroller is used to deduce the sign intended. The output of which is a serial text output in real time The complete hardware working is explained in the next section.

3.2. Approach 2: Using camera as sensor

In this second approach, the gesture/sign by the user is deduced using image processing and computer vision algorithm on a camera sensor. The output of this system us also a text output.

To make this system, we employed the OpenCV library, and in an attempt to pre-process the video frames, we do basic color adjustment and frame axis orientation. To extract the hand from the background, we use the Background Extraction technique to easily recognize the signs made. Using contour extraction and matching we are easily able to identify the unique signals made by the hand. A few preprocessing results are in Fig. 5. The result of which is an alphabet/text which given serially to the system forms a complete sentence.

The result from either of the two approaches as breifly described above is then given as an input to the Text To Speech (TTS) System, which converts the gestures to speech.

4. HARDWARE DESIGN

Approach 1 described in the previous section involves hardware support which works on the principle that the resistance of a piece of metal increases, when it undergoes strain (stretching), in the direction of the strain.

Very similar phenomenon to a piece of wire that is pulled, that will get longer. The diameter will also reduce slightly. Both these physical changes will cause the electrical resistance of the piece of wire to increase. This very small change in resistance can be measured, commonly using several wires arranged in a Wheatstone bridge, and correlated to the stress or force on the material to which the wire is bonded.

As seen in the schematic Fig. 2, instead of a wheat stone bridge to measure the change in the resistance, we use change in voltage to calculate the effect of bending (which tends to give us the same result). Change in the voltage is measured through the digital pins of a microcontroller (Arduino Mega in this case). The rate of change of the voltage is proportional to the rate of bending. The voltage calculated varies in positive or negative direction depending on the direction of the bending.

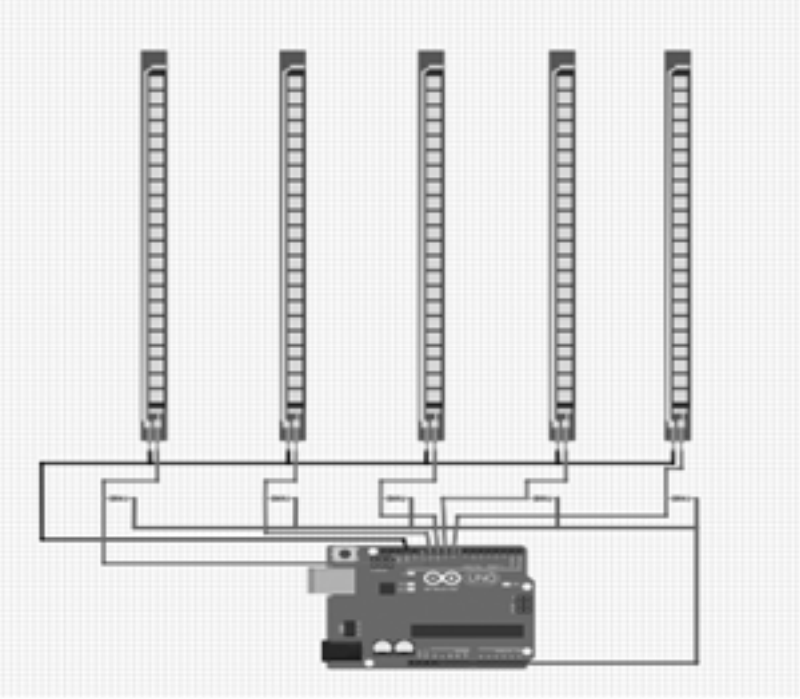

Figure 2: Schematic of the proposed design of gesture glove with flex sensor

Figure 3: A screenshot of the developed text to speech module

A certain threshold is set, which if exceeded by the input voltage gives us a signal that the finger (flex sensor) is bent (This threshold depends on the user capability of folding his/her finger). A certain pattern of finger (flex sensor) movement is recorded in a group of 5 sensors and compared to a stored pattern. This is further given to a text to speech engine serially where it gets converted to an audio signal.

5. SOFTWARE DESIGN

We aim at converting the text that we received from the various flex sensors via the micro-controller to an audio form. This can be carried out by the Text to Speech technique. Here we use the Java Library file in JDK environment on NetBeans IDE to perform this. The library used is freeTTS which is an open source library used for this.

The lib/freetts.jar contains a main entry point that allows a user to interactively control the FreeTTS synthesizer. When invoked with no arguments, FreeTTS will read text from the command line and convert the text to speech as shown in Fig. 3. FreeTTS can also be used to convert text from a file to speech. It

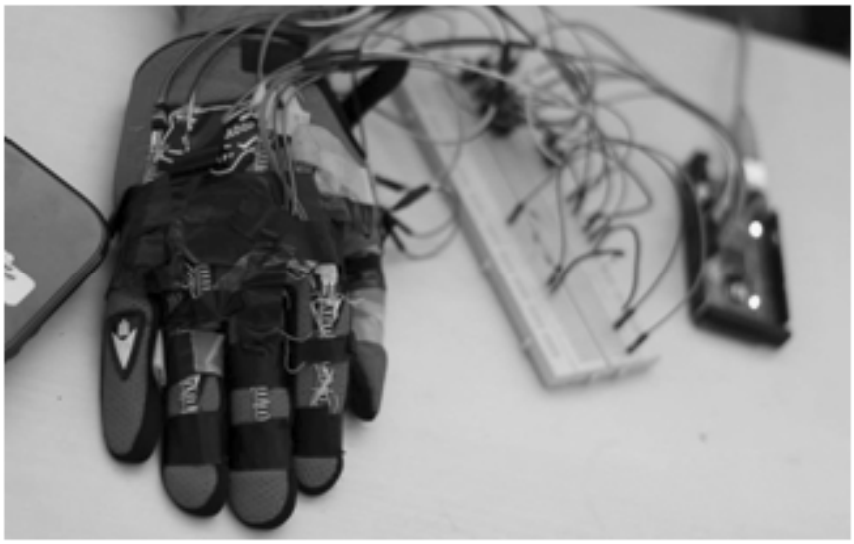

Figure 4: The gesture glove in action

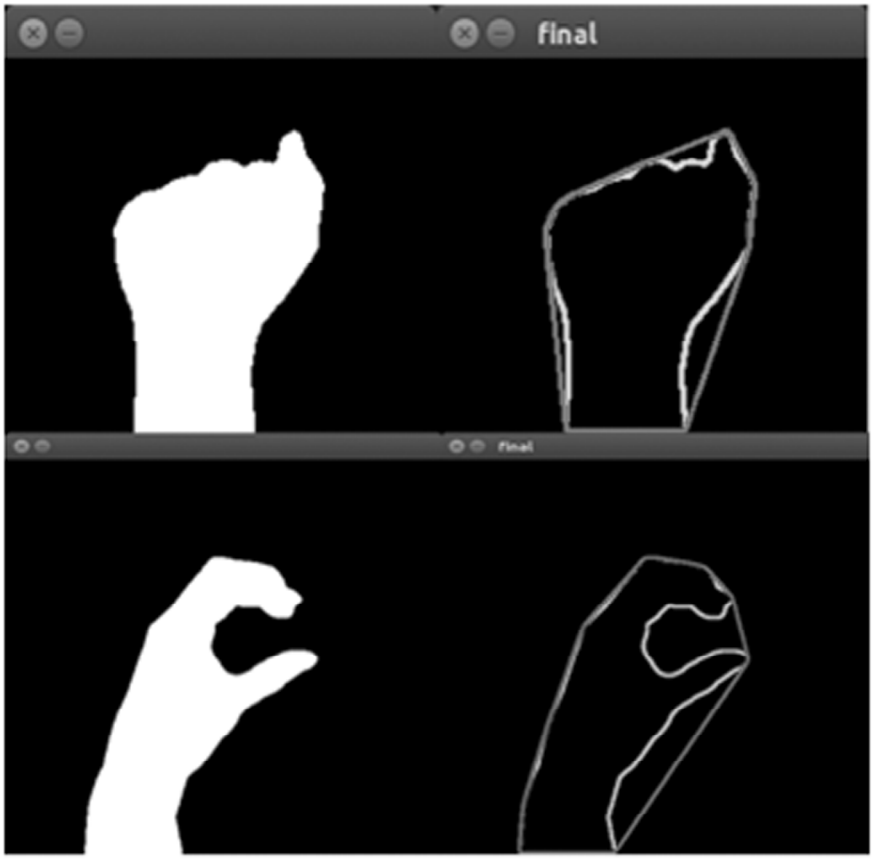

Figure 5: A few video frame preprocessing stage results

includes options that allow you to redirect the audio to file, as well as a number of metrics and debugging options. FreeTTS includes a number of unit and regression tests. The unit tests verify that critical routines are working properly. The regression tests verify that the output of FreeTTS matches what is expected.

6. FUTURE WORK AND CONCLUSION

This paper delineates a novel approach to aid people with speech disability. The work was encouraged after coming across the divergence between people who understand sign language and those who do not. In future work we plan to include a much faster system, and also incorporate autocomplete for familiar phrases.

REFERENCES

[1] "Census Of India: Disabled Population". *Censusindia.gov.in*. N.p. 2011. Web. 19 Jan. 2016. Available: http://censusindia.gov.in/Census_And_You/disabled_population.aspx

[2] "World Report on Disability". Washington: WHO Press, 2011. Print. Available: http://whqlibdoc.who.int/publications/2011/9789240685215_eng.pdf

[3] "Disability Data and Statistics, Monitoring and Evaluation: The Way Forward", Paris: United Nations, 2015. Web. Available: http://www.un.org/disabilities/documents/egm2014/EGM_FINAL_08102014.pdf

[4] "Augmentative and Alternate Communication". Wikipedia. Web. Available: http://en.wikipedia.org/wiki/Augmentative_and_alternative_communication

[5] Rekha, K., Latha, B. ;"Mobile Translation System from Speech Language to HandMotion Language", 2014 International Conference on Intelligent Computing Applications (ICICA) Publication Year: 2014, Page(s): 411–415

[6] Sarji, D.K.; "HandTalk: Assistive Technology for the Deaf", Computer Volume: 41, Issue: 7 Publication Year: 2008, Page(s): 84-86

[7] HaoyunXue, Shengfeng Qin, "Mobile motion gesture design for deaf people" 17th International Conference on Automation and Computing (ICAC), 2011. Publication Year: 2011, Page(s): 46–50

[8] Ohene-Djan, J. Shipsey, R.; "E- Subtitles: Emotional Subtitles as a Technology to Assistthe Deaf and Hearing Impaired when Learning from Television and Film". Sixth International Conference on Advanced Learning Technologies, 2006. Publication Year: 2006, Page(s): 464-466

[9] Taie, M.A. NasrEldin, K.M. ; ElHelw, M.; "ITS navigation and live timetables for the blind based on RFID robotic localization algorithms and Zig Bee broadcasting", Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on Publication Year: 2013 , Page(s): 139–144.

[10] El Hayek, H. Nacouzi, J. Kassem, A., Hamad, M., El-Murr, S. ;"Sign to letter translator system using a hand glove", e-Technologies and Networks for Development (ICeND), 2014 Third International Conference on Publication Year: 2014 , Page(s): 146 – 150

[11] Edwards, A.D.N., "Speech synthesis in augmentative communication", Systems and Applications of Man-Machine Interaction UsingSpeech I/O, IEEE Colloquium on Publication Year: 1991, Page(s): 1-4.

[12] Wang Jian, Han Zhi-qiang, Gao Li, Shen Sheng-bao, Han Zhi-yan,;"Speech visualization simulation research for deaf-mute"; Computer Application and System Modeling (ICCASM), 2010 International Conference on Volume: 7 Publication Year: 2010 , Page(s): 618-621

[13] Soltani F., Eskandari F., Golestan, S.;"Developing a Gesture-Based Game for Deaf/Mute People Using Microsoft Kinect"; Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on Publication Year: 2012, Page(s): 491–495.

[14] Verma, H.V. Aggarwal, E. ; Chandra, S.; "Communication System between Deaf-Blind People and Non-Disabled People Using Body-Braille and Infrared Communication"Image Information Processing (ICIIP), 2013 IEEE Second International Conference on Publication Year: 2013 , Page(s): 96–100.

[15] "UW Augcomm: Features - Rate Enhancement". *Depts.washington.edu*. N.p., 2011. Web. 19 Nov. 2015. Available: http://depts.washington.edu/augcomm/02_features/04d_rateenhance.htm