International Journal of u- and e- Service, Science and Technology
Vol.8, No.1 (2015), pp.131-142
http://dx.doi.org/10.14257/ijunesst.2015.8.1.12

American Sign Language Translation through Sensory Glove; SignSpeak

JanFizza Bukhari, Maryam Rehman, Saman Ishtiaq Malik, Awais M. Kamboh and Ahmad Salman

Department of Electrical Engineering
School of Electrical Engineering & Computer Science
National University of Sciences & Technology, Islamabad, Pakistan
09bicsejbokhari@seecs.edu.pk

Abstract

To make a communication bridge, a highly accurate, cost effective and an independent glove was designed for deaf/mute people to enable them to communicate. The glove translates the sign language gestures into speech according to the American Sign Language Standard. The glove contained flex and contact sensors to detect the movements of the fingers and bending of the palm. In addition an accelerometer was built in the glove to measure the acceleration produced by the changing positions of the hand. Principal Component Analysis (PCA) was used to train the glove into recognizing various gestures, and later classify the gestures into alphabets in real time. The glove then established a Bluetooth link with an Android phone, which was used to display the received letters and words and convert the text into speech. The glove was found to have an accuracy of 92%.

Keywords: *deaf, mute, glove, Euclidian, ASL, alphabet, acquisition, PCA*

1. Introduction

The deaf/mute people make up 72 million of the world's population according to a report published by the World Federation of the Deaf [1]. These people learn sign language to communicate. Unfortunately, most of the average people don't understand their gestures and thus are unable to identify what they are trying to say. This paper is concerned with the solution to help those people having speech disability to have normal conversations in their daily lives.

In this paper, different approaches of gesture recognition are discussed, design of a hand glove for gesture recognition into speech is proposed and the development phases of a complete, independent prototype of sensory glove are elaborated.

Although some data gloves are available in the market but they are used for gaming and other virtual reality applications and there is no such complete system available in the market for the translation of American Sign Language gestures into speech. However, research is being made to devise some portable, efficient and highly accurate system for the translation of standard sign language gestures through a hand glove.

There are two main approaches to gesture recognition: a machine vision based approach which consists of taking the input through a single/set of cameras [2] or a haptic based approach which consists of using a sensory device to take in physical values for processing [3].

1.1 Previous Work

Real-Time American Sign Language Recognition Using Desk and Wearable Computer Based Video using Machine Vision Based Approach [2] represented two vision based SLR systems using hidden Markov models and input through color based tracking. Both

ISSN: 2005-4246 IJUNESST
Copyright ⓒ 2015 SERSC

the systems used a skin color matching algorithm for tracking the hand. The first system had an accuracy of 92% and the second system had an accuracy of 98%. Initially, the image was scanned until an appropriate color pixel was found. Neighbors were checked to find similar color pixels with the result of an outline of the hand.

A boosted classifier tree for hand shape detection [4] provided with an approach to detect the presence of human hands in an image and classifying the hand shape. The location of hands was determined by using a boosted tree structure cascade of classifiers to recognize the shape in grey scale images. For figuring out the exact shape, the data was clustered into similar shapes that contained some variation for the classifier to generalize. The AdaBoost and FloatBoost algorithms were used to find collection of weak classifiers. The k-mediod clustering was applied on the training data to assess the shape similarity according to the distance. Experiments resulted with a 99.8% success rate for hand detection and a 97.4% success at classification.

Visual tracking using depth data [5] used 3D depth data to track hands in cluttered environments. Color and depth information was captured simultaneously using the same optical axis in real time. The speed of capturing depth and color images was around 30 frames per second. The potential fields of possible hand or face contours were determined by the help of 3 algorithms. The first one was based on calculating Manhattan distances and finding the transform to get potential fields. To differentiate between sides of an edge, k-components based potential fields with weights was used. Finally, basin of attraction was used to determine if there was a match. The algorithms were tested on ten people over a period of two months, with a lot of background clutter, and came out with good results.

Hand gesture recognition using Kinect [6] which has sensors to capture both RGB and depth data is Microsoft's Kinect. This system was based on the open NI framework to extract depth data from the Kinect sensor in order to distinguish the user's hand from the background. Accuracies of up to 99% had been recorded while using this system.

A multi-class pattern recognition system for practical finger spelling translation [7], haptic based approach, demonstrated a portable system for recognizing 26 hand shapes of the American Sign Language alphabet using a glove device. Other than alphabets, 'space' and 'enter' were added to the alphabet in order to identify full sentences. The main component of system was the 'Accele Glove' that measured finger position against the gravitational vector. MEMS dual axis accelerometers were attached to fingers, which gave digital output and thus no A/D converter was necessary. Training of the algorithms was done on a PC and disconnected afterwards. The data obtained from sensors was fed into the microcontroller which recognized the letter and sent out the ASCII to a voice synthesizer to speak out the word or sentence. The location of dual axis sensors was on the middle join of each finger to give a measurement proportional to the displacement of mass with respect from the rest position. A duty cycle of train of pulses of 1kHz was used to read the positions.

1.2 Available Gloves in the Market

Different Cyber gloves that were available in the market similar to our construction are: Cyber Glove II by CyberGlove Systems, 5DT Data Glove 14 Ultra and Enable Talk.

Cyber Glove II [8] has a comfortable design with construction of stretchable fabric but the basic CyberGlove II data glove includes one motion capture glove with batteries, a battery charger, and a USB/wireless technology adaptor. The glove comes in two variations: 18-sensor data glove (Price: $11,900.00 USD.) and 22-sensor data glove (Price: $16,795.00 USD.) but these are not complete systems for gesture translation into speech however CyberGlove motion capture system has been used in digital prototype evaluation, virtual reality biomechanics, and animation.

 Copyright ⓒ 2015 SERSC

Similarly 5DT Data Glove 14 Ultra [9] has been designed to satisfy the stringent requirements of modern Motion Capture but it is not a complete independent glove specifically for American Sign Language gestures translation into speech.

Enable Talk [10] project was done by a team, "QuadSquad" from Ukraine, which came first in a Software Design Competition of the 2012 Microsoft Imagine Cup. The system consisted of two parts: hardware part (glove) and a software part, which was developed as a mobile application for Windows 7/8. Bluetooth was used for communication between mobile and glove. The speech was output through Microsoft Speech API and Bing API. EnableTalk team pointed out that Windows Phone 7 didn't allow developer access to the Bluetooth stack and the current version actually runs on windows mobile. They have quoted the price of prototype as $150 but have not discussed its accuracy.

The aim of SignSpeak was to design a user friendly glove that would translate sign language gestures into speech with high level of accuracy for recognizing gestures, should be able to function with limited memory/power constraints and should be able to distinguish between noise and a valid gesture. It should be part of a complete system in which the glove wirelessly communicates with a remote device and should be cheaper than gloves already available in the market.

2. Methodology

To design a glove that would enable deaf and mute people to communicate by translating their sign language gestures into speech according to the American Sign Language, different modules were: glove design, data acquisition system, feature extraction, feature matching, wireless link and android application.

2.1. Glove Design Module

In the designing phase of SignSpeak, selection of appropriate sensors and their location was finalized. Three types of sensors were basically required to detect bending movement of fingers, contacts when two fingers are in contact and acceleration of hand.

2.2. Selection of Appropriate Sensors

Flex sensors of 2.2 inches were used for measuring the bending movements. By noting the resistance of sensor, measurement of bent was done. The sensor was used with another resistance to form a voltage divider, which divides VIN by a ratio determined by the two resistances. The value of resistance used varies for every individual sensor.

To detect a contact, a simpler logic of using conductive plates connected to the input voltage through pull-up resistors was used. Whenever any conductive plate connected to 0V or ground was touched to positive plate, a contact was detected. Hence, whenever one finger was in contact with the other, value of contact sensor (which was initially 1 due to pull-ups) for that particular finger became zero.

To detect and measure the acceleration of hand ADXL 345 was used in I2C mode. Specifically, ADXL 345 was used to recognize the alphabet Z and to distinguish between alphabet I and J. By placing all these sensors on a glove at appropriate locations, a data glove as an input device to our main controller was made.

Copyright ⓒ 2015 SERSC

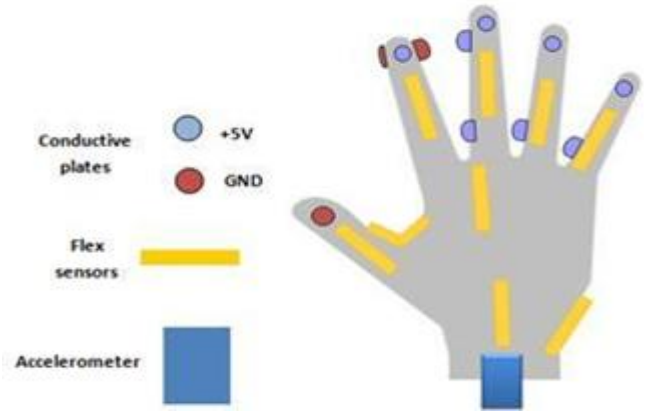


Figure 1. Location of Sensors

2.3. Location of Sensors

To make an ordinary glove a data glove, location of sensors mattered a lot. Since American Sign Language was followed, all the sensors were located such that using less number of sensors, gestures could be easily recognized with maximum degree of freedom of a hand. A total of 21 sensors were used, out of which 9 were flex sensors, 11 were contact sensors and one for measuring acceleration. Location of these sensors is shown in Figure 1.

Referring to Figure 4, flex sensor 8 was placed to distinguish between alphabet A and E, flex sensor 0 was there to distinguish between the gesture of alphabet G and H whereas flex sensor 6 was placed to distinguish between P and Q. Similarly, contact sensors 3, 5 and 6 were placed to recognize alphabet T, N and M respectively. Contact sensor 1 was used to distinguish between alphabet U & V and the GND on the back of index finger was used to detect alphabet R.

2.4. Data Acquisition Module

To acquire multiple samples of gestures for the purpose of extracting important features and training algorithm, a Data Acquisition (DAQ) system was setup, which was able to capture data from the flex and contact sensors. The DAQ system contained PXI module, PXI Controller, PXI Chassis and a Software.

2.5. Hardware used

At least 9 analog voltage channels for flex sensors and at least 10 digital channels for contact sensors were required. To enhance the performance and accuracy of the system, signal conditioning for accelerometer readings was needed. Baud rate was set at 9600 and sampling rate at 500 samples per second. The device used was: NI PCI-6250. SCB-68A I/O block was used for interfacing signals to plug in DAQ devices with 68 pin connectors. The cable used was NI SHC68-68-EPM which was specially designed to work with NI X Series and M Series data acquisition (DAQ) devices, and included separate digital and analog sections, individually shielded twisted pairs for analog inputs, individually shielded analog outputs, and twisted pairs for critical digital I/O.

The NI PXI 8330 module was used to connect an external controller (our desktop computer) to a PXI chassis. It was paired with a PCI 8330 in desktop computer with a MXI cable running between the two. It contained a MXI-3 Multi-System Extension Interface for PCI.

 Copyright ⓒ 2015 SERSC

2.6. Software Interface

LabVIEW System Design Software was used to make custom VIs that allowed acquiring and analyzing the measurement of data. Data samples at a rate of 100 samples / second was acquired for a total of 2.5 seconds each, and saved it in data logs in a folder.

2.7. Feature Extraction

For the purpose of classification, data was needed to be arranged according to the difference in features of each gesture. A variant of machine learning algorithm for Adaptive Boosting, named Gentle AdaBoost was tried which was a well-known method to build ensembles of classifiers with very good performance. The concept was to weigh the data instead of (randomly) sampling it and discarding it.

Although gestures were classified very efficiently and accurately using this algorithm, but the amount of complexity needed for its implementation was not required in case of SignSpeak, and a simpler algorithm would also work well with less number of computations. So, Principle Component Analysis was applied to real time data obtained from sensors.

A total of 26 gestures were trained, with 20 recordings of each consisting of 250 samples each. Taking averages, final master set was a 520 X 17 matrix where 17 was the number of dimensions. Principle Component Analysis was used for classification and feature extraction to reduce dimensionality while preserving as much class discriminatory information as possible. Input to PCA was a data matrix n x p where n corresponds to observations i.e. 520 for SignSpeak training set, and p corresponds to no. of variables which is 17 (as is the number of sensors). Output of PCA was p x p matrix of Principle Component Coefficient and loadings. MATLAB built-in functions and commands were used to determine the PCA coefficients. For every real time input, PCA vector was multiplied with the input to get input PCA vector as in Eq. (1)

These coefficients along with corresponding means of the alphabets were used to classify the real time input using the Euclidean distance formula.

Input_PCA = input * PCA_Vector (1)
A_PCA = A_PCA – input_PCA. (2)
B_PCA = B_PCA – input_PCA. (3)
……..

Y_PCA = Y_PCA – input_PCA. (4)
Z_PCA = Z_PCA – input_PCA. (5)
……..

Norm (A_PCA). (6)
Norm (B_PCA). (7)
……..

Norm (Y_PCA). (8)
Norm (Z_PCA). (9)

2.8. Feature Matching

The (p x p) matrix of PCA coefficients stored in our microcontroller was used for matching with the real time input to determine a valid gesture. Once the program is started, Arduino takes continuous input from all the sensors attached on the glove until a stable input is received. The condition for receiving a stable input is:

Copyright ⓒ 2015 SERSC

Minimum - x <= Maximum <= Minimum + x
The input is in binary format, and to convert it into a voltage value, the following formula was used: Voltage = bits * (3.3V/1023)

The real time inputs are processed with PCA coefficients and their corresponding Euclidean distances are found. The gesture with the most nominal distance is considered as a most valid gesture. After determining a valid gesture, accelerometer is activated to record any possible acceleration.

2.9. Wireless (Bluetooth) Link and Android Application

C-05 (Bluetooth to Serial Port Module) was used because of its ease of availability in Islamabad and its low cost. Service Discovery Protocol (SDP) was used to allow devices to discover what services each other support, and what parameters to use to connect them. Each service was identified by a universally unique identifier.

The default setting of serial port of HC05 was: baud 38400 and our serial communication in the Arduino was happening at a baud rate of 9600, thus baud rate of Bluetooth module of SignSpeak was needed to be matched. To change it, a command was entered through the AT mode on the module. The module was set as Master mode.

2.10. Transmitting Data from Arduino to PC

Before transmission, it was necessary to pair the device with the computer by adding the device from device manager (Windows). At pairing, LED connected to PI09 of Bluetooth module blinked very fast and slowed down when a connection was established. When the module is transmitting data, LED connected to PI08 blinks twice at slow intervals to indicate data transfer. At the Arduino side, serial print functions were used to send text to the PC terminal.

2.11. Transmitting Data from Arduino to Android Device

The connection steps of Arduino and Bluetooth module remained the same. For receiving the data on an android device, an application was developed for this purpose which used Bluetooth API. The application was written in Java using Eclipse, by modifying the Bluetooth Chat sample code from Android. Two APIs Bluetooth API (for scanning of other Bluetooth devices and paired Bluetooth connections) and Text to Speech API (for conversion of text into speech featuring more than 30+ languages and implementation of different features such as voice, speed and pitch) were used. The main flow of Bluetooth link is shown in Figure 2.

3. Results and Discussion

Initial calibration of sensors showed that three of the sensors had almost a constant increase voltage recorded when fingers in stretched straight position for 5 seconds, similarly a constant (but comparatively increased) voltage when in closed fully bent position for 5 seconds, and when the fingers were moved from straight to fully bent, there was a similar pattern of voltage rising in all 3 sensors. Plus, their starting and ending voltages matched with the previous ones, thus verifying the consistency and repeatability of the result (Figure 3).

Some of the graphs obtained during data acquisition for some alphabets are shown in Figure 5 while some screenshots of working video of the glove are given in Figure 6.

The machine vision based approach [2] could have varied results because of the difference in people's skin color, and the lack of ability to distinguish hands from face. It also requires the user to wear clothes with full sleeves to cover arms. Plus, lighting effects could adversely affect this method.

 Copyright ⓒ 2015 SERSC

However, SignSpeak provides a solution independent of any camera, which is equally accurate in light as in the dark. Similarly, for hand shape detection, a boosted classifier tree approach [4] could give inaccurate results when, for both training and test databases, hand images have fairly simple and similar backgrounds whereas SignSpeak approach is not based on images. While gesture recognition using Kinect [6] requires the setup of Kinect that has sensors to capture both RGB and depth data for gesture translation whereas SignSpeak is independent glove which does not require any such device.

Measuring how reproducible experimental measurements of SignSpeak were, the Euclidean distance of each set was noted from the average of 23 training sets initially. When the percentage of these differences was taken, it came out to be 85%. Later on, the accuracy was improved by fixing one of the flex sensor placed at the wrist bent. When more training sets were taken, accuracy was found to be around 92% for an untrained user.

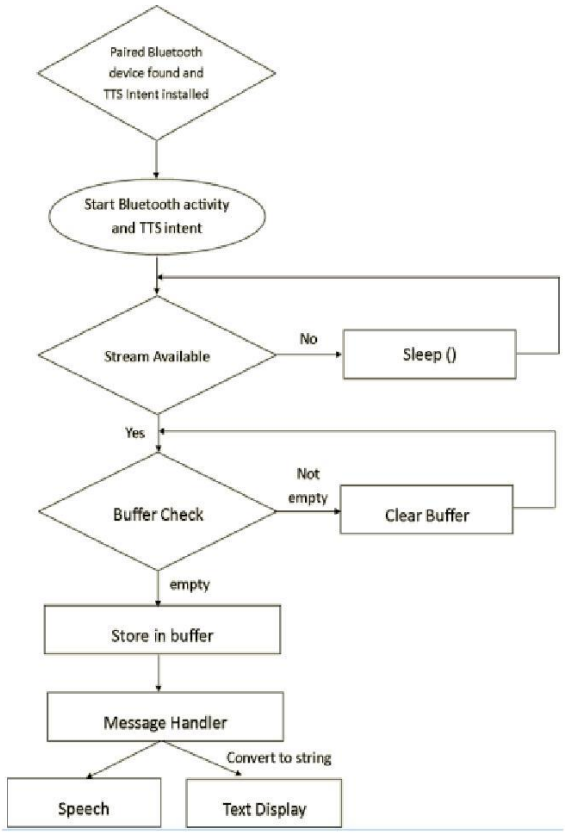


Figure 2. Flow of Bluetooth Link

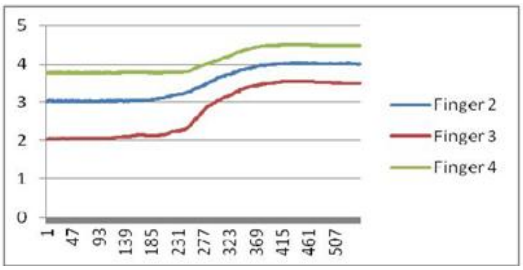


Figure 3. The Movement of Fingers from Straight to Fully Bent Position

Copyright ⓒ 2015 SERSC

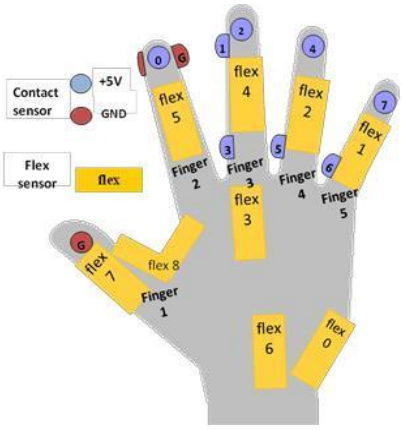


Figure 4. Labeling of Flex and Contact Sensors on the Glove

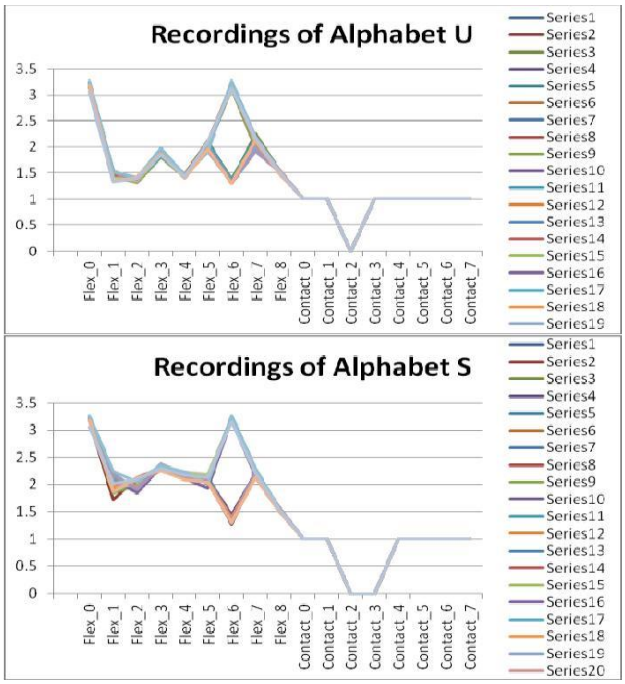
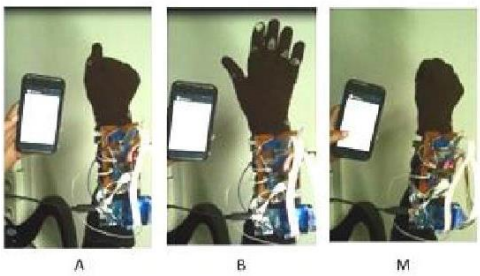


Figure 5. Data Acquisition for Alphabets



 Copyright ⓒ 2015 SERSC

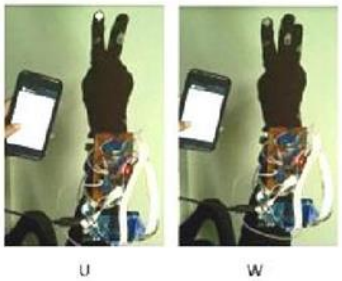


Figure 6. Screenshots of the Working Glove Showing Various Alphabets

4. Conclusion

To make a communication bridge, a highly accurate, cost effective and an independent glove was designed for deaf and mute people. The glove is capable of translating their sign language gestures into speech according to the American Sign Language Standard through android phone. SignSpeak focuses the translation of gestures of the alphabets; however, sensor placing is such that a few words can be added in the library of gestures without any change in the design of the glove. Comparing with other approaches, SignSpeak uses Principle Component Analysis to classify the real time input data for feature extraction. Recognizing gesture of the similar/ non-similar alphabets, SignSpeak is found highly accurate.

This paper focuses the translation of sign language through the glove; however, other possible applications of the Glove could be: its use for interaction between the human hand and virtual reality environments, robotics and for art and entertainment.

References

[1] Murray, J.: Deaf women and girls not able to enjoy basic human rights, A submission by the World Federation of the Deaf on the half day of general discussion on women and girls with disabilities, Geneva (2013)
[2] Starner, T., Weaver, J., Pentland, A.: Real-Time American Sign Language Recognition using Desk and Wearable Computer Based Video, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, No. 12 (1998)
[3] Parvini, F., McLeod, D., Shahabi, C., Navai, B., Zali, B., Ghandeharizadeh, S.: An Approach to Glove-Based Gesture Recognition, California 90089-0781 (2009)
[4] Ong, J., Bowden, R.: A Boosted Classifier Tree for Hand Shape Detection, FGR' 04 Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition, ISBN: 0-7695-2122-3, pp 889-894, USA (2004)
[5] Nanda, H., Fujimura, K.: Computer Vision and Pattern Recognition Workshop, IEEE Xplore, 0-7695-2158-4/04 IEEE (2004)
[6] Li, Y.: Hand Gesture Recognition using Kinect, Thesis submitted to the Faculty of the J. B. Speed School of Engineering of the University of Louisville (2012)
[7] Hernández-Rebollar, J.L., Lindeman, R. W., Kyriakopoulos, N.: A Multi-Class Pattern Recognition System for Practical Finger Spelling Translation, Fourth IEEE Int'l Conference on Multimodal Interfaces (ICMI'02), Pittsburgh, USA, pp 185-190 (2002)
[8] Yazadi, F., Cyber Glove Systems, LLC, 2157 O'Toole Avenue, Suite L San Jose, CA 95131, http://cyberglovesystems.com/ ?q=products/cyberglove-ii/overview
[9] Anonymous: 5DT 5th dimension technologies, http://www.5dt.com/products/ pdataglove5u.html
[10] Anonymous : Microsoft Imagine Cup, Sydney, http://www.microsoft.com/enus/news/ Press/2012/Jul12/07- 10ImagineCup12PR.aspx (2012)

Copyright ⓒ 2015 SERSC

Authors

JanFizza Bukhari, she received her BE degree in Information and Communication Systems Engineering from School of Electrical Engineering & Computer Science, National University of Sciences and Technology, Pakistan in 2013. Her research interests include Wireless Ad-hoc Networks, Digital Signal Processing and Software Defined Networking. Engr. JanFizza is a registered member of Pakistan Engineering Council. Currently, she is serving SciFlair Research Lab NUST Pakistan as Research Assistant contributing to research and development in the area of Ad-hoc networks. She has also worked on a project "Android Automation" with experienced professionals of SMART Research Lab, NUST Pakistan in 2012.

Maryam Rehman, she received her BE degree from NUST School of Electrical Engineering and Computer Science Pakistan in 2013. During her studies, she has been assisting the professors for Numerical Analysis and Digital Logic Design as Teaching Assistant. Maryam is awarded Rector's Gold Medal for Best Project in BICSE for the session 2013. Presently, she is working as a Software Developer at Aurora Solutions, Pakistan.

Saman Ishtiaq Malik, she graduated in 2013 from SEECS NUST as Information and Communication Systems Engineer. She is pursuing her MS in Innovative Technologies in Education from NUST School of Electrical Engineering and Computer Science Pakistan. Recently, she has won the ILM Apps Challenge; a series of interventions aimed at creating platforms for e/m-learning. Saman is working as QA Associate at Agnitus, Pakistan since 2013.

Awais M. Kamboh, he received his BSE degree in Electrical Engineering from College of E&ME, NUST in 2003, his MSE degree in Electrical Engineering Systems from University of Michigan, Ann Arbor, USA in 2005 and his Ph.D. in 2010 from Michigan State University, East Lansing, USA. His research interests include Biomedical Signal Processing, VLSI and low-power ASIC Design. In addition to taking on various research and teaching assignments, he has worked as ASIC Design Engineer at Evigia Systems, Ann Arbor, USA. Dr. Awais is an HEC approved PhD supervisor, and a member of the IEEE and the PEC. His research articles have been published in several international journals and conferences. He received the best paper award in IEEE BioCaS 2010. His research on high-density neural data recorders was also featured in IEEE Spectrum blog.

Dr. Awais is the director of Analog and Mixed Signal Research Group at SEECS.

 Copyright ⓒ 2015 SERSC

Ahmad Salman, he received his PhD degree in Computer Science from University of Manchester, United Kingdom in 2011. His research interests include Machine Learning and Speech Information Processing. Dr. Ahmad is serving NUST as Assistant Professor since September 2012. He has also worked for SUPARCO Pakistan as Assistant Manager during June 2003 – June 2004.

Copyright ⓒ 2015 SERSC

Copyright ⓒ 2015 SERSC