Sign Language Translation in Urdu/Hindi Through Microsoft Kinect

Sohail¹, M Y Khan¹, N A Siddiqui² and Kiran³
¹ Faculty of Computing, Mohammad Ali Jinnah University, Pakistan
² Department of Computer Sciences, Sukkur IBA University, Pakistan
³ Department of Computer Sciences, Technical University Berlin, Germany

E-mail: sohailahmedhanfi@gmail.com
E-mail: yaseen.khan@jinnah.edu

Abstract. Communication is the foremost and basic necessity of a person. The ability to speak without any problem is a boon; nevertheless, woefully we have often seen people around us who are the victim of speech or hearing impairments. Thus, the ‘Sign Language’ (SL) appears as an alternate standard language that is globally understood and adopted by the deaf community. Though SL is a communication tool, but in practice, we still see most of the normal people do not understand the SL properly; thus, it is again a problem in communication between a speech/hearing impaired and normal persons. In this regard, many research attempts have been made to resolve this problem via wearable technology and other different paradigms of computational solutions. However, almost all of them focused on English or western languages, which provide a trivial resolution for the people of the Indian subcontinent. Thus, in this paper, we propose a solution to address this issue by employing Kinect Motion Sensor for detecting the signs (in SL) through hand movements and gestures and translating them in Urdu/Hindi language, which is readable and audible to normal persons. The proposed system is developed on a feature-based model; hence, we stored signs in a dictionary/training set for the development of the baseline system; however, the users can customize the system by adding their sign patterns in the future.

1. Introduction
Humans are exceptional amongst all living beings with the ability to speak, which permits vocalizing in order to communicate among themselves understandably. Thus a human who can speak, narrowly without flaw, can be characterized as a normal being. Regrettably, some people are not capable of speaking due to major deficiency, particularly with problems in the articulately or hearing organs, which makes them impaired to speak or hear [1] Hence, society has developed Sign Language (SL) as a medium of communication with subjects. The SL utilizes the gestures and movements of hands and fingers to precise the thought of subjects. The syntax and morphology of the SL play an important role, and the standard components (i.e., lip movements, facial expression, etc.) can produce a significant change in the meaning of a sign [2] However, these SLs differ from region-to-region as we see the change in natural languages worldwide [3]. The SL, although, offers a conversational medium among the subjects, but for the rest of the people, the situation is still the same.

In the recent era, we have seen computational treatments/solutions for overcoming the issue. The scientists and linguistics appraise various methods of Sign Language Translation (SLT) in
this field. In the same regard, this paper focuses on the development of a tool-based approach for automated SLT. The underlying theme of the proposed work (as depicted in fig 1) is to employ the Microsoft Kinect sensor for identifying gestures in SL and then translating them into words that are supposed not only for displaying on-screen but also audible to normal people.

1.1. Research Goals and Objectives
The goal of this research is to develop an automated Sign language translator that uses the data provided by the Microsoft Kinect XBox 360 TM sensor. Kinect works based on human joints as the X, Y and Z coordinates of these joints are moved Kinect catch them, and we can handle the results of motion in back end programming. An input sign given by the user is captured by Kinect camera and infrared sensor and translator provide the exact word in English and Urdu/Hindi language. One can easily comprehend the purpose of observing figure 1.

The work is done for this paper targets Indo-Pakistan SL as an input and Urdu/Hindi languages for the output language. Thus, with the two output languages (i.e., Urdu/Hindi which are mutually intelligible [4], and considered as the 3rd most spoken language w.r.t native speakers [5]), we address the problem for Indian subcontinent who have almost 7.4% of the disabled population in Pakistan [6] and 6.3% of the total population in India[7], respectively, are affected by hearing or auditory loss/impairment; which grosses around ~ 99.8 million people (all together for both countries).

The other goals of this research work presented hereunto:
- Design a user-friendly interface so that users can easily understand and run the system.
- System should be capable of performing efficiently and produce results in run time so that the user can communicate in a real-time environment.
- Make user capable of storing its signs in sign dictionary for both languages, i.e., English and Urdu/Hindi. All of these features make a deaf person able to communicate with a normal person, which helps him to overcome his deficiency though it serves the society with positive outcomes.

2. Related Work
In Pakistan, Boltay Haath System [8] is a well-known gesture recognition system that is computer-based that converts signs produced by deaf into Urdu speech. Their technique works
by taking input as an incoming sign and compares it with a dataset of signs/gestures. They have got an accuracy level of nearly 70%.

Fuzzy classifier, which is another approach proposed by Kausar et al. [9], it used a classification algorithm based on fuzzy logic to recognize signs/gestures shown by the deaf. In their algorithm, an angle is measured for the classification of gestures between fingers and their joints. Accuracy of their work is 95%.

Haseeb et al. [10] built an inverse system named PSLIM for the deaf community in another way. In their system, the focus is on reducing the communication gap between the hearing impaired and hearing person. They convert sign/gesture of PSL by listening to the audio. They have achieved 78%.

Ali [11] came up with another solution that helped enhance communication between normal and deaf person. Their proposed system takes input as both text and image of a sign and translates it to another form. Their system is tested upon the Urdu alphabets of PSL using the Haar classifier with an RGB camera. However, not only the accuracy of the system is missing, but the nature of the dataset is too much complex for experimentation.

Khan et al. [12] successfully Converted PSL signs/gestures into Urdu language with the help of image processing. Their system provides easiness to allow deaf as a novice to learn computer programming and software engineering etc. It also proposed a comprehensive frame-work and used a subset of C++. The used RGB camera to capture 500 images of 37 letters of Urdu alphabets. Their technique works with skin region filtering using image processing. After that, they applied discrete wavelet transformation (DWT) by extracting features of interest. It is the first technique that is software-based without any hardware that is data glove etc. The presentation of their form is in textual form. They have selected 226 images out of 500 for training and 74% for testing and achieved accuracy 86%.

Sami et al. [13] recognized Urdu alphabets in Pakistan Sign Language (PSL). In their solution, computer vision technology was used for the recognition of PSL Urdu. They have found a close match between the input image and the image dataset. The accuracy of their work is achieved 75%.

Although a great deal of related work for SLT is mentioned above, the proposed work is uniquely based upon Microsoft Kinect 360 and allows the user to store his signs in sign dictionary to work for the Urdu and Hindi language.

3. Methodology

The proposed solution is divided into four main modules, as shown in the block diagram (see figure 2). The details of these modules are covered in the sub-sections 3.1 – 3.4, respectively.

Figure 2: The architecture of the sign language translation system.
3.1. Human Gesture Detection
The Motion Detection System uses Microsoft Kinect for the Xbox 360TM device along with its SDK 1.7 to detect the signs of the deaf person; these signs consist of gestures that are performed by any deaf person in front of the Kinect. The Kinect consists of three basic parts that collaborate to sense any activity and produce the real image on the screen: an RGB color VGA video camera, a depth sensor, and a multi-array microphone [14]. The camera spots the facial expressions, body type, and color mechanism i.e., red, green, and blue. Its pixel resolution and frame rate are 640×480 and 30 fps, respectively. This aids in facial and body identification.

The depth sensor has a monochrome CMOS sensor and infrared projector that facilitates to form the 3D imagery all over the room. It also calculates the remoteness of each point of the user’s body by transmitting invisible near-infrared light and measuring its ‘time of flight’ after it replicates the objects. These components can sense and trace 48 diverse points on each user’s body and repeats 30 times every second. A famous algorithm of ‘Dynamic time warping (DTW)’ has also been used to accurately identify the sign performed by an individual user while he is changing his position or multiple users are present in front of sensors.

3.2. Dynamic Time Wrapping (DTW)
This algorithm was presented in the 60s; it is modal to identify closeness between two instances [15], which may shift in time or speed. For example, the algorithm will recognize the pattern of a walking person regardless of whether in one instance the individual is walking in slow steps and if in another video he or she is walking fast, or regardless of whether there are increasing or decreasing velocities. By utilizing DTW, the system will have the capacity to locate an ideal match between two given instances (i.e., signs) with certain confinements. The instances are wrapped non-linearly in the time measurement to decide the measure of their likeliness In this research project, DTW has been utilized along with the input of Kinect sensor for motion/sign detection to identify the user even if he is changing the position more frequently and with different sign executions speeds.

3.3. Sign to Text Conversion
The detected signs are compared with the signs already stored in the dictionary each time the Kinect sensor captures a frame. As the sign is marched, the sign to text converter results in the meaning of the sign in either English or Urdu/Hindi language, whatever the user selects language mode. A flat file has been used to store gesture coordinates and their corresponding meaning in English and Urdu/Hindi language.

3.4. Text to Speech Converter
For English text-to-speech conversion, Microsoft’s text to speech synthesizer library for .NET is used that takes English text as input and creates audio stream as output. It is also referred to as text-to-speech (TTS). A synthesizer must present an extensive analysis and processing to correctly alter a string of characters into an audio stream that seems just as the words would be uttered.

For Urdu/Hindi text-to-speech conversion, a database of own voices is created against every gesture in the dictionary, so when any gesture is detected, the system plays the voice of related word from the database in the Urdu language.

Users can capture and store his signs in the dictionary by using a simple gesture capturing system in the proposed solution. User is required to set himself in front of the Kinect sensor and see its skeleton; click on the capture button and write the name of sign that he wants to
store the gesture takes 32 frames and store the sign performed in 3 seconds against the name of
gesture given by the user.

3.5. Flow of Activities
The Activity starts with Deaf/Signer, who stays in front of the device, he will perform sign
actions in the form of sign language Motions Detection system detects these sign by Using
Kinect. The Sign to text system will check the sign language according to the Singer input; if
signs are correct according to the dictionary of gestures, it will convert it into the English or
Urdu text. The system will convert that text into English or Urdu audible form with the help
of a speech synthesizer, and Listener will have the respectable voice of those signs which he/she
performs.

4. Result
A total of 24 signs have been stored in the sign dictionary, which are thoroughly tested with 10
subjects, alongside people without speech/hearing impairments. Since, it is a baseline idea, and
a lot of sign words have to be employed, therefore, authors found the results with the currently
stored signs are quite encouraging. The system successfully takes signs as input and convert it
into English Urdu, and Hindi text and followed by conversion of the audio.

The authors also add that the paper has a limited space, therefore, a sample of most frequently
experimented sign words are presented in table 1. Further, with the same reason of limit, we
only present the coordinate values for the word ‘NO’ in the table 2, to adjust the width of the table, the respective values are rounded off to 4 decimal places, however during the experiment the values were upto 15 decimal places.

Table 1: A sample of word list used in the proposed system.

| English | Urdu | Hindi | Transliteration |
|---------|------|-------|-----------------|
| I       | مُحمَد | मुझे | /mʊhɜːd/        |
| Help    | مَدد   | मदद | /m@d/          |
| Need    | چاہیے  | चाहिये | /tʃæhjɪ:/      |
| Hello   | हेलो  | हेŶो | /he:lo:/        |
| Thanks  | شُكریہ | शुक्रिया | /ʃʊkɜːriːya:/ |
| No      | نہیں  | नहीं | /n@hi/         |
| Why     | کیسے | क्यों | /kɪjʊ/         |

Table 2: Coordinate values for word ‘NO’ in Indo-Pak SL, captured in Microsoft Kinect.

| Microsoft Kinect Coordinate Values for ‘No’ Sign |
|-----------------------------------------------|
| (-0.7769, -1.9738, -0.7099, -1.6615, -0.7178, -0.9049, 0.8706, -0.659, 0.7453, -0.1675, 0.7491, -0.0312) |
| (-0.7768, -1.9742, -0.7904, -1.6612, -0.7178, -0.9047, 0.8707, -0.6605, 0.7525, -0.1512, 0.7487, -0.03) |
| (-0.7758, -1.974, -0.7897, -1.6604, -0.717, -0.9048, 0.8628, -0.6561, 0.7494, -0.1721, 0.7505, -0.0422) |
| (-0.7754, -1.9742, -0.7896, -1.6601, -0.7165, -0.9044, 0.8657, -0.6573, 0.7477, -0.1706, 0.7504, -0.0406) |
| (-0.774, -1.9741, -0.7881, -1.6601, -0.7161, -0.9044, 0.867, -0.6585, 0.7478, -0.1704, 0.7509, -0.0372) |
| (-0.7732, -1.9741, -0.7874, -1.6605, -0.7153, -0.9048, 0.8641, -0.6583, 0.7508, -0.1735, 0.7513, -0.037) |
| (-0.7732, -1.9744, -0.7874, -1.6604, -0.7149, -0.9045, 0.8659, -0.6605, 0.7502, -0.1733, 0.7518, -0.0392) |
| (-0.7731, -1.9746, -0.7873, -1.6611, -0.7148, -0.9051, 0.8667, -0.6617, 0.7505, -0.1735, 0.7535, -0.0439) |
| (-0.773, -1.9756, -0.7874, -1.6615, -0.7147, -0.9052, 0.8683, -0.6626, 0.7494, -0.1725, 0.7535, -0.0414) |
| (-0.7724, -1.9745, -0.7867, -1.6608, -0.7143, -0.905, 0.8678, -0.663, 0.7485, -0.1732, 0.7525, -0.0469) |
| (-0.7724, -1.9749, -0.7866, -1.6609, -0.714, -0.9046, 0.8687, -0.6643, 0.7476, -0.171, 0.7531, -0.0488) |
| (-0.7727, -1.9753, -0.7864, -1.6612, -0.7139, -0.9049, 0.8688, -0.6649, 0.75, -0.1712, 0.7528, -0.0472) |
| (-0.7724, -1.9747, -0.786, -1.6601, -0.7137, -0.9044, 0.8679, -0.6652, 0.7493, -0.171, 0.7511, -0.0441) |
| (-0.7719, -1.9745, -0.7867, -1.6602, -0.7137, -0.9046, 0.8684, -0.6662, 0.7492, -0.1715, 0.7514, -0.0451) |
| (-0.7725, -1.9757, -0.7872, -1.6609, -0.714, -0.9046, 0.8686, -0.6668, 0.7507, -0.1707, 0.7522, -0.0424) |
| (-0.773, -1.9774, -0.7875, -1.6628, -0.7141, -0.9061, 0.8702, -0.6688, 0.7511, -0.1719, 0.7541, -0.0499) |
| (-0.773, -1.9781, -0.788, -1.6623, -0.7137, -0.9062, 0.87, -0.6692, 0.7508, -0.1724, 0.7543, -0.053) |

5. Conclusion
We conclude that our proposed system, i.e. for the translation of Indo-Pak sign language in respective speech/audios, shows positive results w.r.t available signs in user-defined dictionary. The system is also capable to grow the dictionary through the addition of signs. Although the data is limited for this research work, but it proves the potential of the proposed methodology. This research work provides an opportunity for the development of sign language translator in different languages. Alongside expanding the sign-dictionary, we aim to provide this solution for portable devices in the future so that deaf people can carry the system with them and get more benefit from the system.
References

[1] Deafness and hearing loss - world health organization https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss accessed: 2020-04-10

[2] Elliott E A and Jacobs A M 2013 Frontiers in psychology 4 115

[3] Hassan B, Farooq M S, Abid A and Sabir N 2015 VFAST Transactions on Software Engineering 9 1–6

[4] Khan M Y and Nizami M S 2020 2020 International Conference on Information Science and Communication Technology (ICISCT) (IEEE)

[5] Khan M Y, Emauddin S M and Junejo K N 2017 2017 IEEE 11th International Conference on Semantic Computing (ICSC) pp 242–249

[6] Niazi F With sign language, i am equal - the news https://www.thenews.com.pk/amp/152779-With-Sign-Language-I-am-Equal accessed: 2020-04-10

[7] Varshney S e t al. 2016 Indian journal of otology 22 73

[8] Alvi A, Azhar M, Usman M, Muntaz S, Rafiq S, Rehman R and Ahmed I 2007 International Journal of Computer, Electrical, Automation, Control and Information Engineering, World Academy Of Science, Engineering and Technology 1

[9] Kausar S, Javed M Y and Sohail S 2008 Proceedings of the 8th conference on Signal processing, computational geometry and artificial vision (World Scientific and Engineering Academy and Society (WSEAS)) pp 101–105

[10] Haseeb A A, Ilyas A J and Ilyas A 2012 Speech translation into pakistan sign language

[11] Ali S A Detection of urdu sign language using harr algorithms

[12] Khan N S, Abid A, Abid K, Farooq U, Farooq M S and Jameel H 2015 Speak pakistan: Challenges in developing pakistan sign language using information technology

[13] Sami M, Ahmed H, Wahid A, Siraj U, Ahmed F, Shahid S and Shah S I A 2014 2014 International Conference on Robotics and Emerging Allied Technologies in Engineering (iCREATE) (IEEE) pp 200–204

[14] Zhang Z 2012 IEEE MultiMedia 19 4–10

[15] 2007 Dynamic Time Warping (Berlin, Heidelberg: Springer Berlin Heidelberg) pp 69–84 ISBN 978-3-540-74048-3 URL https://doi.org/10.1007/978-3-540-74048-3