A Communication Platform Between Bangla and Sign Language

Rhythm Shahriar, A.G.M. Zaman
Computer Science Department
American International University - Bangladesh
Dhaka, Bangladesh
{rythmshariar@gmail.com}

Tanvir Ahmed, Saqib Mahtab Khan, H.M. Maruf
Computer Science Department
American International University - Bangladesh
Dhaka, Bangladesh

***Abstract*—Hearing and speaking impairments are very crippling handicaps for human beings. People affected by such impairments often use sign language to communicate, which is very different from spoken and written language, often leading to difficulties in communication when a translation medium is not readily available. This leads to heavy struggle in the daily lives of the affected people, especially in Bangladesh. A simple, accurate and portable translation medium is necessary to address this problem, which would allow two-way communication between a normal person and an impaired person. Speech recognition and conversion to sign language is a key part of such a medium, especially since Bangla is a rather challenging language to work with. In this paper, we have described the development of a smartphone application, which provides a two-way communication platform, which employs a speech-to-sign language system and a text-to-speech system. Testing of the application shows satisfactory results, with an average accuracy of 84.71% in various test environments.**

Keywords—bangla; bangla sign language; bangla speech to sign; bangla text to speech; sign language; cdd sign language; cmu sphinx; communication platform; deaf communication; bridge; medium;

I. Introduction

Sign language is a visual language that uses a system of various body movements as a means of communication. Much like natural languages, there are many different forms of sign languages used in different countries around the world. Sign language is used by the people with hearing and speaking impairments to communicate with each other. In most cases, sign languages are never written down but are used only for direct visual communication, such as interpersonal communication and video broadcasts. Also, the sign languages are usually grammatically different than their natural language counterpart. Many people with hearing and speaking impairments may not know the natural language at all, while many people without such impairments would not know how to communicate in sign language. These differences make it very difficult for these two types of people to communicate effectively with each other without a translator. Human translators are inefficient and there are not enough of them to make sure every personal communication can take place between the two communities of people.

The proposed research topic is a system for digitally converting Bangla speech to Bangla sign language, to enable effective communication between the two communities. This system will recognize Bangla speech and convert it to text, then translate it into Bangla sign language with the help of a developed algorithm. Another system will be provided for converting text to speech, to ensure two-way communication. These two systems, when used together, will help a hearing or speech impaired person to communicate with a regular person very effectively without a human translator. This would be beneficial to society, as these physically disadvantaged people can now communicate and integrate themselves into the everyday life.

II. Related Works

This literature review aims to review, compare and critique the research done so far on how Bangla text and speech can be translated into Bangla sign language, and some similar research with similar goals in other countries. Efforts are being made around the world to digitally translate natural languages into their sign language counterparts, and vice versa. Although several attempts have been made to develop a system to visually recognize Bangla sign language and convert it to Bangla text [1] [2], very limited research has been done to accomplish the reverse. Therefore, it is necessary to study relevant research from other countries, in addition to all the research available from Bangladesh, to potentially discover new ideas.

K. Datta, B. Sarkar, C. Datta et al develops a word-based translation system framework for Bangla text to Bangla sign language. It processes the input text with a Bangla dictionary, arranges the words in the correct order for Bangla sign language, then displays the appropriate video file associated with each word [3]. In another paper published by the same group, details are explained about how the arrangement of words would take place. An important thing to note are the main syntax rules in this translator: if a noun comes after an adjective in Bangla text, those two words would be swapped with each other; any interrogative word (e.g. what, why, where, etc.) would be placed at the end of the sentence; connective words (e.g. but, and, otherwise etc.) would be dropped from the sentence; certain words would be replaced with two associated words and display two video clips to convey the word (e.g.: hotel = eat + money). These are some basic grammar rules in Bangla sign language [4].

Similar studies around the world follow similar methods to translate their languages. Some of them involve translating English text to their sign language, which could be used with

978-1-5386-2175-2/17/$31.00 ©2017 IEEE

another system which translates another language to English (e.g. Google Translate). A. Haseeb and A. Illyas from Pakistan develops a system to translate English speech to Pakistani sign language [5]. In the parsing and translation part of this system, the English sentences are analyzed twice, first analyzing for phrase structures, and then analyzing for grammatical relations. Using this analysis, the nature of the sentence is determined, and words and expressions are displayed accordingly as sign language clips [5].

There has been a few research efforts to convert Bangla speech to text. Some of these involve third-party interfaces as one of the steps for conversion. S. Sultana, M. Akhand, P. Das and M. Rahman uses SAPI, a speech interface from Microsoft, in their efforts to recognize Bangla speech. Although SAPI itself does not have the ability to directly convert speech to text for Bangla language, the team uses the English language has a middleware to manage SAPI. They store an XML file containing Bangla words written with English characters, which can be matched with speech input based on pronunciation matching. The output is displayed using text written in Avro Bangla software. [6] Another team, consisting of N. Hossain, D. Shahreen and T. Kazi, uses SAPI and similar techniques to display the text output instantly and using Bijoy Bangla software [7]. Some other studies have other different methods of recognizing Bangla speech. M. Rahman, F. Khan and A. Bhuiyan use an audio segmentation and separation technique to classify speech patterns and derive words from the speech using feature extraction. The words are classified by syllables and length, which are then used for feature extraction by Mel Frequency Cepstral Coefficients (MFCC) [8]. A. Hasnat, J. Mowla and M.

Khan employ the Hidden Markov Model to recognize Bangla speech, and breaks down the sounds into phonetics instead of words. Then they show the difference between recognizing isolated speech and recognizing continuous speech [9]. M. Rahman and F. Khatun works on converting only isolated Bangla words in audio format to text, by first recording the words and storing them in wav files. Then they classify the words according to their number of syllables, convert the audio signals into digital form for feature extraction by MFCC and employ direct Euclidean distance measurement technique for speech recognition. The accuracy of their system ranged from 79% to 96% depending on the speaker [10].

Some more efforts to convert speech to the text were attempted with the use of the CMU Sphinx speech recognition system. CMU Sphinx is a group of speech recognition systems, which consists of various speech recognizers and an acoustic model trainer. These systems are used to recognize and convert speech to text in various languages, including Bangla. CMU Sphinx systems are a good choice of frame-works for converting Bangla speech to text, as they are very flexible, different packages can be used with different programming languages and they have many features [11].

S. Chowdhury implements a speech recognition system for Bangla using the CMU Sphinx system, where they first select a small domain of phrases to work on and then note some accompanying information such as age, gender, dialect, environment and device details for working with the data. They also use the CRBLP pronunciation dictionary, a language model file with "arpa" format among other components to facilitate the speech recognition. The resulting accuracy of the experiments conducted in this research are rather high, ranging from 70% to 90%, and it shows that this is a good start for converting Bangla speech to text [12].

Considering all the research covered in this literature review, the best approach for developing a Bangla to sign language translator would be a simple and effective one: taking a speech input through the mic, converting the speech to text using the CMU Sphinx framework, then reading and analyzing the text, and then simply rearranging the words according to Bangla sign language grammar, turning them into tokens, and spelling out the missing words in between. Furthermore, developments can be made for Bangla language to be easily machine translated to Bangla sign language, just like any other language.

The main research question for the objectives would be: How to create a simple and effective translation tool for two-way communication between a deaf/dumb person and a normal person in Bangla language. To answer the main research question, the sub-questions to be answered are: which tool or interface to use for converting Bangla speech to Bangla text, how to create a large database for storing the tokens, how to convert the Bangla text into sign language, how will the data be collected, which platform is best for using the system as an application, how much output data is sufficient for testing, and whether or not the algorithm works sufficiently.

III. Methodology

The objective of this research is to find a solution, and develop a system for two-way communication, the main goals were recognized at first as structured objectives, then the potential solutions and details were explored as the research progressed. The data collected and used in this research were all experimental data, derived from experiments on the software. The block diagram illustrated in figure 1 shows the overview of the whole communication system.

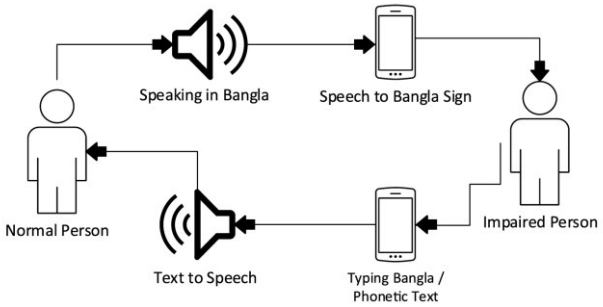

Figure 1: Overview of the communication system

IV. Implementation

The CMU Sphinx speech recognition system is set up using macOS. Therefore, files and commands compatible with

macOS are used in this process. The following libraries were used to set up the environment:

i. sphinxbase-5prealpha

ii. sphinxtrain-5prealpha

iii. pocketsphinx-5prealpha

A. Bangla Speech to Sign Language

The speech to sign language system is partly facilitated by the CMU Sphinx speech recognition system. The Sphinx is used to define the Bangla phonetics, words and grammar, which would recognize the speech input and convert it to phonetics text. The phonetics text is converted into full words using our algorithm, and then finally the words are matched with the sign language database to show the sign language images on the screen as shown in figure 2.

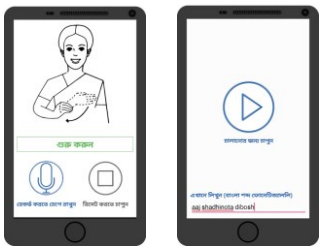

Figure 2: Bangla speech to sign and text to speech

- **Bangla Dictionary:** A scope of words is selected from the Bangla language dictionary. These words are broken into syllables and stored in a phonetic dictionary, such as a .dic file. All the words that are to be recognized, are stored here.

- **Bangla Language Model:** Some sentences are built using the previously selected scope, and they are saved as the Bangla language model file. The file is uploaded to the imtool website, the .lm file is retrieved from the result and converted to a .lm.bin file. A filler file is used to reduce noises and a .phone file is used to store all the syllables in the language.

- **Bangla Acoustic Model:** The recorded files are in a .wav format with the following properties:
Sampling rate: 16000 Hz, Bit rate: 16 bits, Channel: Mono. Media normalized to reduce noise. Additional files are used to indicate the audio files which are used for training and the strings of sentences spoken by trainers.

- **Training the Model:** The directory which contains all the relevant files is opened and the *setup* command is used to begin the training process, which generates some files to be configured as recommended. Finally, the *run* command is used to complete the building process.

- **Speech to Sign Display Conversion Process:** First, the recorded speech is broken down into phonetics text using the language model. Then, the phonetics text is converted into Bangla words using an algorithm in figure 3.

```
1. READ (UserVoice)
2. INITIALISE (RecognizedText)
3. SPLIT (RecognizedText)
4. INITIALISE (RecognizedArray)
5. INITIALISE (ResultArray)

6. for i = 0 to RecognizedArray.length
7.     for j = 0 to DatabaseArray.length
8.         if RecognizedArray[i] = DatabaseArray[j]
9.             push into ResultArray

10. for i = 0 to ResultArray.length
11.     display ResultArray[i]
```

Figure 3: Voice to Sign Conversion Algorithm

Then, the Bangla words are matched against the list of words in the Sign Language database, and if any word is detected, the corresponding images are shown on the display as Bangla Sign Language output. Figure 4 demonstrate whole process of speech to Bangla sign language conversion process. The user interface would have the option to play and pause the images and record speech for conversion.

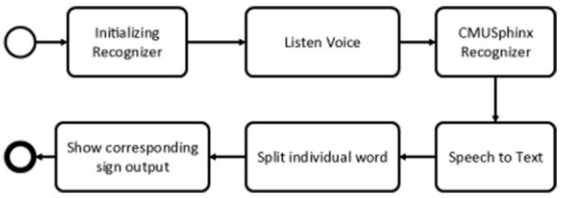

Figure 4: State diagram of speech to sign language

B. Bangla Text to Speech

The text to speech conversion of Bangla language is done using the Google Translate Server. Text is entered and sent to the server, which converts it to speech and sends back the corresponding audio stream as figure 2.

- **Text Input and Processing:** At first, a text input is entered into the text box. This input text is processed before sending to the Google translation server. At first, the length of the input string is recorded. Then, the words are split by detecting the spaces between the words. Finally, a "+" symbol is added in between each word and prepared for sending to the server.

- **HTTP request to Google Translate Server:** The processed string is sent to the Google Translate Server as a HTTP request. The Google Translate Server converts the text to speech and generates a media file, ready to be streamed directly.

- **Streaming from Google Translate Server:** Upon pressing the "Play" button, the generated media is played directly from the Google Translate Server. The stream can be paused and resumed and replayed. The whole process of the text to speech is illustrated in figure 5.

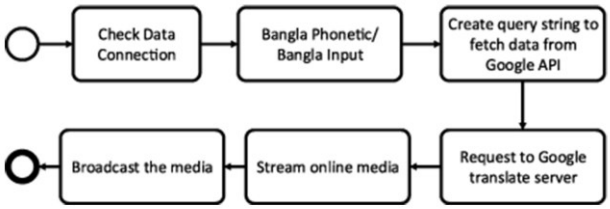

Figure 5: State diagram of text to speech conversion

V. Result Analysis

The system was tested by varying the environment and the devices. The tests were conducted in three different environments: lab room, living room, and a noisy area, using two devices at each environment: An Android smartphone and a simulator (Android Studio) figure 6 shows the performance analysis.

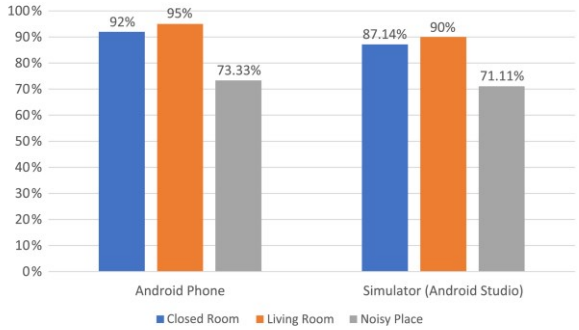

Figure 6: Performance analysis of the system

From the results, we find that the average accuracy is 86.67% with an Android smartphone, and 82.75% with the simulator. Also, from an environment standpoint, the system is 89.57% accurate in the lab room, 92.5% accurate in the living room and 72.22% in the noisy area.

Comparing the different speech recognition techniques such as MFCC [8] [10] and SAPI [6] [7]. A third-party interface was considered the easiest and most efficient method to achieve the desired result. The main drawback of SAPI is that it cannot be trained, and American accent has to be used for the technique to work properly, as well as converting to English language, which is an additional and unnecessary step. Furthermore, CMU Sphinx method produces more consistent and accurate results than the MFCC method.

VI. Conclusion and Future Work

The system described in this paper can be used for two-way communication. The system shows satisfactory results. It might be difficult for a speaking or hearing-impaired person to understand static images of sign language. Therefore, we would also like to work on replacing the image output of the sign language with animation or video output, for easier understanding during communication.

The text to speech part of the application is just a small part of the system we implemented to ensure two-way communication, but we would like to expand on that system in the future to turn it into a sign language to speech model, by using techniques like image processing.

The system can be made more useful quickly using crowd-sourcing methods. People can help expand the database of trained words, and increase the accuracy of speech recognition. This is an efficient way of entering a massive amount of different speech data in a small amount of time.

References

[1] N. N. Choudhury, G. Kayas et al., "Automatic recognition of bangla sign language," Ph.D. dissertation, BRAC University, 2012.

[2] A. Jarman, S. Arshad, N. Alam et al., "An automated bengali sign language recognition system based on fingertip finder algorithm," International Journal of Electronics & Informatics, 2015.

[3] K. Datta, B. Sarkar, C. Datta, D. Sarkar, S. J. Dutta, I. D. Roy, A. Paul, J. U. Molla, and A. Paul, "Interface between bangla text and sign language."

[4] B. Sarkar, K. Datta, C. Datta, D. Sarkar, S. J. Dutta, I. D. Roy, A. Paul, J. U. Molla, and A. Paul, "A translator for bangla text to sign language," in 2009 Annual IEEE India Conference. IEEE, 2009, pp. 1–4.

[5] A. A. Haseeb and A. Ilyas, "Speech translation into pakistan sign language," 2012.

[6] S. Sultana, M. Akhand, P. K. Das, and M. H. Rahman, "Bangla speech-to-text conversion using sapi," in Computer and Communication Engineering (ICCCE), 2012 International Conference on. IEEE, 2012, pp. 385–390.

[7] N. Hossain, D. Shareen, and T. Kazi, "Instant bangla speech to text conversion."

[8] M. M. Rahman, M. F. Khan, and M. A.-A. Bhuiyan, "Continuous bangla speech segmentation, classification and feature extraction," IJCSI Interna-tional Journal of Computer Science Issues, vol. 9, no. 2, pp. 1694–0814, 2012.

[9] M. Hasnat, J. Mowla, M. Khan et al., "Isolated and continuous bangla speech recognition: implementation, performance and application perspec-tive," 2007.

[10] M. Rahman and F. Khatun, "Development of isolated speech recognition system for bangla words," vol. 6, no. 1. Daffodil International University Journal of Science and Technology, 2011.

[11] N. Shmyrev, "CMUSphinx Open Source Speech Recognition", CMUSphinx Open Source Speech Recognition. [Online]. Available: http://cmusphinx.sourceforge.net. [Accessed: 25- Oct- 2016].

[12] S. A. Chowdhury, "Implementation of speech recognition system for bangla," Ph.D. dissertation, BRAC University, 2010.