2021International Conference on Computing, Communication, and Intelligent Systems(ICCCIS)

Speech To Indian Sign Language (ISL) Translation System

Pankaj Sonawane
Computer Engineering
D. J. Sanghvi College of Engineering
Mumbai, Maharashtra
pankaj.sonawane@djsce.ac.in

Karan Shah
Computer Engineering
D. J. Sanghvi College of Engineering
Mumbai, Maharashtra
karan12011998@gmail.com

Parth Patel
Computer Engineering
D. J. Sanghvi College of Engineering
Mumbai, Maharashtra
parthptl2297@gmail.com

Shikhar Shah
Computer Engineering
D. J. Sanghvi College of Engineering
Mumbai, Maharashtra
sb.shah14@gmail.com

Jay Shah
Computer Engineering
D. J. Sanghvi College of Engineering
Mumbai, Maharashtra
jaysh1397@gmail.com

2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS) | 978-1-7281-8529-3/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCIS51004.2021.9397097

***Abstract*—The deaf and dumb population across the Indian subcontinent predominantly communicate with each other and with other people using the Indian Sign Language (ISL). ISL employs signs consisting of hand gestures, facial expressions or body postures to convey the desired message and emotion. It is a full-fledged natural language with its own grammar and lexicon. In order to lessen this significant communication gap between the hearing and speaking impaired community and the normal population, there is a need for translation systems. We propose an end-to-end human interface framework that is capable of recognizing and interpreting spoken language and then act out the corresponding ISL gestures to facilitate a very convenient, real time form of conversation between the disabled community and the rest of the population. We made use of the Microsoft Xbox Kinect 360s depth sensing and motion capturing abilities to capture motion data for all the different ISL gestures and then used Unity3D to set up all the animations and then finally bundle everything into an Android application.**

Keywords—Indian Sign Language; Sign language translation; 3D Modeling; Motion capture

I. Introduction

According to the 2011 census, 63 million people in India suffer from a from of significant hearing disability which adds up to 6.3% of the entire population of the country [2]. Of these 63 million people, it is estimated that at least 76-89% of them have no knowledge of Indian Sign Language [2]. The reasons for this comparatively low literacy rate are believed to be: 1) Lack of ISL interpreters 2) Unavailability of ISL tool 3) Lack of research on ISL. Speech and hearing impaired people face a lot of communication issues due to their limited ability. This poses quite a problem for the growth and development of disabled people. The most pervasive solution to this problem is the use of sign language for the purpose of communication. While the usual spoken languages are used by the hearing population to express thoughts and ideas, a sign language like ISL is a visual language based on gestures and action serves as the primary means of communication for hard hearing individuals.

ISL is the primary sign language used by the deaf community in the Indian subcontinent. The linguistic studies of ISL began in 1978. Initial linguistic work finds that ISL is a language in its own right. It was found that 75% signs are same across the region. ISL includes both word level gestures as well as fingerspelling. Fingerspelling is used for the words for which there is no direct sign language translation, which the signer does not know the gestures of and for emphasis or clarification of a particular word. Such words can be formed letter by letter using fingerspelling. The fingerspelling gestures could be both static as well as dynamic. The signs in ISL are either one-handed or two-handed consisting of three main parts: manual features which include hand gestures, non-manual features which include facial expressions and body posture which contribute to the sign expressed, and fingerspelling. In the later decade of 2000s, research began on ISL recognition systems. The key problems faced by researchers in such systems are as follows: 1) Maximum of more than 2700 gestures in ISL are dynamic which make it difficult for recognition. 2) The identification of pronouns, tense, prepositions and other parts of speech is a challenge when it comes to ISL to English conversion. 3) There are differences in the dialects spoken across India. 4) There is no genuine and completed ISL database available. Researchers who want to work on the development of such systems need to create their own database of signs.

II. LITERATURE REVIEW

Purushottam Kar et al. [20] in their 2007 paper developed INGIT, a system for translating Hindi strings to Indian Sign Language. It was developed specifically for Railway Inquiry domain. FCG was used to implement the grammar for Hindi. The developed module converts the user input into a thin semantic structure. Unnecessary words are removed by feeding this input to ellipsis resolution. The ISL generator module then generated a suitable ISL-tag structure depending on the type of sentence. A graphical simulation was then generated by a HamNoSys converter. The system was successful for about 60% cases in generating the semantic structures.

Ali et al. [21] developed a domain-specific system in which the input fed had to be English text. The text was converted into ISL text which was further translated into ISL symbols. The architecture of the system had the following components:1) A text translation input module. 2) Tokenizer to break down the sentence into separate words. 3) A ISL symbols repository which was specific to railway inquiries. If a word had no corresponding sign assigned to it in the repository, then it’s synonyms sign was used. 4) All the words were mapped with their corresponding symbols by a purposefully built translator. It also filtered the words to be translated by eliminating the words which were offensive or abusive or did not have any sign stored. 5) An accumulator which accumulated the words in the sequence entered.

Vij et al. [22] developed a 2-phase system of Sign Language Generation. The first phase dealt with preprocessing Hindi 2 Sentences and converting it into ISL grammar. The phase used a combination of Dependency Parser and WordNet for this purpose. Dependency graphs in the Dependency Parser represented words and their relationships between head words and words which modify those heads. In the second phase, HamNoSys was used for converting this grammar into different corresponding Sign Language symbols. The generated symbols are converted into XML tags form using SIGML. The XML tags form is then readable by a 3D rendering software.

MS Anand et al. [3] developed a two-way ISL translation system. In the speech-to-sign module, the input speech was first put through the noise removal submodule. The output was then used as an input for the speech recognizer for decoding the spoken speech into a textual word sequence. A natural language converted the word sequence into a sequence of signs by a rule-based technique. Finally, a sign animation module with text annotation was used for displaying the signs.

In the system generated by Dasgupta et al. [17]. English text was taken as input and it was then converted into the corresponding ISL structure that adhered to the rules of the grammar. Their system comprised of the following key modules: a) Analysing text coupled with parsing Syntax b) Representation using LFG f-structure c) Transfering grammar rules and finally d) Generating proper ISL sentences . Minipar Parser was used to parse the input sentence and the parse tree was used to construct a dependency structure. An f-structure is generated which encoded the grammatical relation of the input sentence. When we say Grammatical relation of the input sentence, we essentially refer to the subject, object and tense of the sentence. We represented this This information as a set of attribute-value pairs. Each attribute corresponded to the actual grammatical symbol name. On applying proper grammar transfer rules, the English f-structure generated was converted to Indian Sign Language f-structure. It needs to be mentioned that evaluating this system is extremely difficult due to the unavailability of a proper, official ISL written orthography.

III. METHODOLGY

UsingMicrosoft Kinect for Xbox 360, we first recorded motion capture data for all the basic Indian Sign Language gestures. The recording process required one of us to physically act out each individual gesture in a welllit room with minimum disturbances in the background. The captured motion capture files were labeled and stored serially in our master database. Simultaneously, in Blender 3D, we mapped out a humanoid mesh on a fbx 3D avatar created by us on Autodesk Maya. The motion capture files were then imported onto the mesh in Blender and the corresponding composite fbx file was generated for each gesture. All the composite fbx files were then finally imported into our Main Scene in Unity where each one of them were individually branched out in the animator hierarchy. Various other parameters like the playback speed, start time, end time, loop time and pose data were set in the animation controller. The Google Speech-to-Text service available natively in Android was then called upon on click of a button which converted spoken input into parseableStrings. The Speech-to-Text service requires network access and hence all the necessary runtime permissions were explicitly provided for via code. The result of the Speech-to-Text API was stored as a composite String. This String was passed as an input to our ISL parser. Our ISL parser separates out all the words in an order as prescribed by the rules of ISL. If any word was found not present in our library, that word was broken down into alphabets and each alphabet was processed independently. We also accounted for slangs and offensive words and they are automatically discarded by our system. We had to make use of multiple threads in the background as the enactment of the gesture took a considerable chunk of the processing time on the main thread. All actions had to be synchronized as well. The word/letter currently being enacted is highlighted so as to ease the process of communication and for visual aesthetics. The Unity package is then converted into an APK file with the help

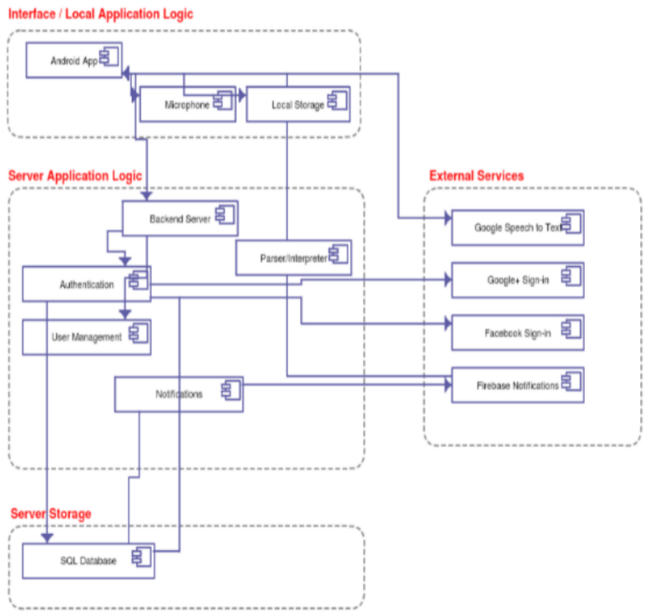

Fig.1: System Architecture

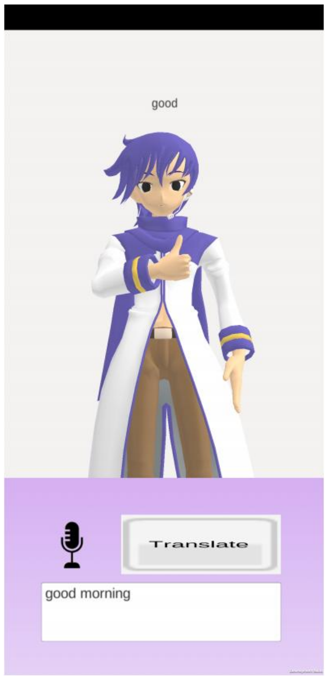

Fig. 2: User Interface

of the Android SDK for Unity 3D which can be shared between mall the supported Android devices or shared on the Google Play store to ensure maximum visibility of the app across the globe.

IV. User Interface

Designing the user interface was always going to be challenging. We adopted a minimal approach in preparing the application layout and kept everything as simple as possible. Our primary aim was to display the 3D character in the 3 centre of the screen so that it occupies the maximum available real estate. The speak and the translate buttons were kept in the bottom half of the screen for better accessibility. We also provided for a text input layout just in case giving a voice input was infeasible. Color Scheme was well thought off and calibrated so as to follow the general Android design guidelines.

We also provided for real time feedback of the current word being enacted by highlighting that part of the sentence. This was particularly useful when the spoken sentence was too long and the ISL gestures did not follow the order in which the words were spoken in the first place. Haptic feedback was provided on button clicks and on successful translation so as to add to the overall user experience. The ability to customize the avatar based on user preferences to add a personal touch can be included as a part of the next iteration.

V. Results

During testing among our peers, we found that our application worked properly in 91% cases over a week of regular usage. A few words were misinterpreted due to the limitations of the speech-to-text engine and hence the required sign was not displayed. In all other scenarios where the words were correctly interpreted, the system seemed to function exceptionally well. When we visited the institution for the deaf and dumb people, testing the application with 100 deaf and dumb subjects. We identified 5 signs that the deaf and dumb people did not understand. These signs were re-enacted and fed back to the system. Also, the experts suggested that a few signs were incomplete for their corresponding phrases and context. Another important factor which needed expert counselling was regarding the pace at which the signs were performed We eventually decided to keep the speed at which our avatar performs signs adjustable/variable as different deaf and dumb people have varying grasping powers. Adjustable speed increases the usability factor of the application as it enhances the ease of sign language translation. Also, the application received an average usability of 7.6 through the feedback form. The feedback form gave us an overall idea about the usability of the application in real world scenario and the ways in which we could improve it.

VI. Discussion

Our paper brings many important points about ISL translation systems in the discussion. A review of the existing literature on such systems reveals a major disparity in their research. Majority of the research projects in this domain have focused on ISL recognition and conversion and only few studies have implemented the translation in the opposite direction. There is a need for more research in development of a sign-language translation system for two main reasons: 1) It can act as an assisting system for deaf people. 2) It can serve as a good basis for the development of gestural human machine interfaces. 2) The lack of annotated or motion ISL data is a key hindrance to research in the area. Researchers who want to work on development of such systems need to create their own database of signs. Accurate ISL gesture motion data can only be generated using sophisticated devices like data gloves or Leap Motion controller which can prove to be an expensive affair. Developing a dictionary including motion data about hand shape and location, facial expressions and palm orientation of the ISL corpus is one possible solution to this problem.

VII. Conclusion and Future Scope

The primary objective of our paper is to emphasize on the significance of translation systems for Indian Sign Language and to develop a working solution which will translate speech to ISL. Major research has been done in other popular languages like American Sign Language while ISL has seen less recognition in this regard. Development of this system will help bridge the communication chasm between hearing and speech impaired community and the able community. Inclusion of all communities leads to the progression of a society as a whole.

There are maximum of more than 2700 dynamic gestures in ISL. We intend to incorporate more and more gestures in our system, as many as we can. There needs to be a continued, prolonged and regular updation of our gesture library in order to make this product truly exhaustive and useful. The main challenge when it comes to ISL to English 4 conversion is to correctly identify nouns, pronouns, tense, prepositions and other parts of speech. Identifying the tone of the sentence and the corresponding emotions need to be better conveyed. A feedback mechanism will ensure that our system will keep getting better. Our goal is an 100% accuracy while translating speech to Indian Sign Language. There are a lot of different dialects spoken all across India. Every state, every province has its own dialect which makes this translation more difficult than ever. Via Machine Learning and a very robust, enhanced Natural Language Processing, provisions for all the various dialects have to be made. Native Language Support is the key for any successful product in India. There is no genuine and completed ISL database available. We plan to incorporate all our ISL gesture data and make it open source. This would serve as a very useful material for the general public as well as the fellow researchers and developers.

Dumb people are usually deprived of normal communication with the general public. The only choices they have are to rely on an interpreter or some form of visual communication. Moreover, an interpreter can not be always available, so through this paper we can help eliminate the dependency on the interpreter. To have a complete understanding of the tone and context of input speech, the system could be extended to integrate the knowledge of facial expressions and body language too. A mobile application and a web application will increase the reach to more people. The system can be made open source so that more and more people can record signs for corresponding words and feed them into the existing data set. Hence, it will increase the size of data set so that more sign language translations for different words can be obtained.

References

[1] Indian Sign Language text generation from English/Hindi Text, Pawan Kumar, Savita Khatri et al. International Journal of Recent Research Aspects (IJRRA) ISSN: 2349-7688, Special Issue: Conscientious and Unimpeachable Technologies 2016, pp. 30-33

[2] National Sample Survey Organization. Disabled persons in India. NSS 58th round (July December 2002) Report no. 485 (58/ 26/ 1).New Delhi: National Sample Survey Organization, Ministry of Statistics and Program Implementation, Government of India, 2003.

[3] M.Suresh Anand, A.Kumaresan, Dr.N.Mohan Kumar, An Integrated Two Way ISL (Indian Sign Language) Translation System A New Approach, Volume 4, No. 2, Jan-Feb 2013 International Journal of Advanced Research in Computer Science.

[4] Hee-Deok Yang, Stan Sclaroff,and Seong-Whan Lee, Sign Language Spotting with a Threshold Model Based on Conditional Random Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 7,pp.1264- 1277, July 2009.

[5] Archana S. Ghotkar, Rucha Khatal , Sanjana Khupase, Surbhi Asati Mithila Hadap, Hand Gesture Recognition for Indian Sign Language, 2012 International Conference on Computer Communication and Informatics (ICCCI -2012), Jan. 10 12, 2012, Coimbatore, INDIA.

[6] Yanan Xu, Yunhai Dai, Review of Hand Gesture Recognition Study and Application, Contemporary Engineering Sciences, Vol. 10, 2017, no. 8, 375 384.

[7] Ananya Choudhury, Anjan Kumar Talukdar, Kandarpa Kumar Sarma, A Conditional Random Field Based Indian Sign Language Recognition System under Complex Background,2014 Fourth International Conference on Communication Systems and Network Technologies.

[8] Gajanan K. Kharate, Archana S. Ghotkar, Hand Segmentation Techniques to Hand Gesture Recognition for Natural Human Computer Interaction.

[9] Dr.C. Chandrasekar, G.T. Shrivakshan, A Comparison of various Edge Detection Techniques used in Image Processing, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 5, No 1, September 2012.

[10] Samta Gupta , Susmita Ghosh Mazumdar, Sobel Edge Detection Algorithm, International Journal of Computer Science and Management Research.

[11] Lei Yang,Xiaoyu Wu, Dewei Zhao, Jun Zhai, Hui Li, An Improved Prewitt Algorithm for edge detection based on noised image, 2011 4th International Congress on Image and Signal Processing.

[12] Shilpa Bhople , Prof. R. R. Itkarkar , Pooja Bhoi, International Journal of Electrical Electronics Computer Science Engineering Volume 2, Issue 3 (June, 2015) — E-ISSN : 2348-2273 — P-ISSN : 2454-1222.

[13] Bag, S.; Sanyal, G., ”An efficient face recognition approach using PCA and minimum distance classifier,” Image Information Processing (ICIIP), 2011 International Conference on , vol., no., pp.1,6, 3-5 Nov. 2011.

[14] G. R. S. Murthy R. S. Jadon, A review of vision based hand gestures recognition”, International Journal of Information Technology and Knowledge Management July-December 2009, Volume 2, No. 2, pp. 405-410.

[15] A. El-Sawah, N. Georganas, E. Petriu, A prototype for 3-D hand tracking and gesture estimation, IEEE Transactions on Instrumentation and Measurement, 57(8):16271636, 2008.

[16] Sergei Azernikov. Sweeping solids on manifolds. In Symposium on Solid and Physical Modeling, pages 249255, 2008.

[17] Tirthankar Dasgupta et al. Prototype Machine Translation System from Text-to-Indian Sign Language.

[18] John Canny. A computational approach to edge detection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, PAMI-8(6):679698, Nov. 1986.

[19] Thomas B. Moeslund. Image and Video Processing. August 2008.

[20] Purushottam Kar, Madhusudan Reddy, Amitabha Mukerjee and Achla M. Raina, “Limited Domain Formulaic Translation from Hindi Strings to Indian Sign Language”, INGIT

[21] S. F. Ali, Gouri Sankar Mishra, A. K. Sahoo, “Domain Bounded English To Indian Sign Language Translation Model”, International Journal of Computer Science and Informatics, Volume 4, Issue 1, Article 6

[22] Paras Vij, Parteek Kumar, “Mapping Hindi Text To Indian sign language with Extension Using Wordnet”, AICTC '16: Proceedings of the International Conference on Advances in Information Communication Technology & ComputingAugust 2016 Article No.: 38 Pages 1–5