

Sign Language Generation - A survey of techniques



Shruti Chikalthankar, Archana Ghotkar

Abstract: Sign language is a visual language that uses body postures and facial expressions. It is generally used by hearing-impaired people as a source of communication. According to the World Health Organization (WHO), around 466 million people (5% of the world population) are with hearing and speech impairment. Normal people generally do not understand this sign language and hence there is a communication gap between hearing-impaired and other people. Different phonemic scripts were developed such as HamNoSys notation that describes sign language using symbols. With the development in the field of artificial intelligence, we are now able to overcome the limitations of communication with people using different languages. Sign language translating system is the one that converts sign to text or speech whereas sign language generating system is the one that converts speech or text to sign language. Sign language generating systems were developed so that normal people can use this system to display signs to hearing-impaired people.

This survey consists of a comparative study of approaches and techniques that are used to generate sign language. We have discussed general architecture and applications of the sign language generating system.

Keywords: HamNoSys, Machine translation, Natural Language Processing, Sign language.

I. INTRODUCTION

Sign language is the primary language of hearing impaired people. It is a completely established language which has its own grammar and lexicon. Unlike acoustically conveyed sound patterns, sign language uses body language and manual communication to convey the thoughts of a person. It is performed by simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions. It is difficult for hearing impaired people to access the information because of their language problem and hence this hinders their normal social life.

Since sign language is the fundamental language of hearing impaired people, many people face difficulty in reading or writing the complex text. Also to communicate with other people, hearing impaired people have to struggle to make the person understand what he/she says.

Or they need a human translator to translate sign language into speech. In crowded areas such as railway platforms, banks, hospitals or theaters, generally hearing impaired people have to suffer for the information.

Like there are different spoken languages in the world, there are different sign languages in different countries. They are developed independently of the spoken language in a particular region. For example, British Sign Language (BSL) and American Sign Language (ASL) are different, even though the spoken language used by normal people of Britain and America is the same.

Indian sign language and Pakistan sign language are similar to Japanese sign language (JSL), Taiwanese sign language (TSL), and Korean sign language (KSL) are similar to each other.

Indian Sign Language Research and Training Centre (ISLRTC) made ISL certified interpreter's lists of various organizations/institutions/colleges/university. This list has been further divided into six zones: North, South, East, West, Central and North-East. From the graph it can be observed that very limited human interpreters are available to assist hearing impaired people of India. [21]

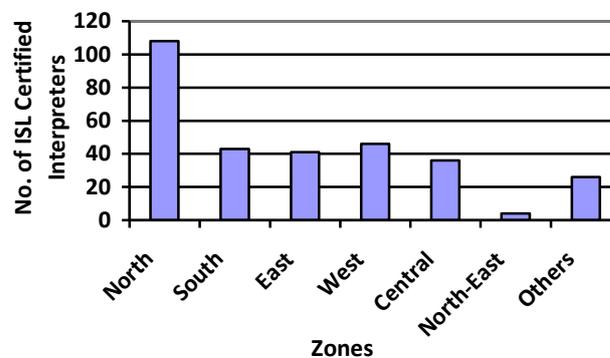


Fig. 1. Zonal Distribution of Certified ISL interpreters in India [21]

II. SIGN LANGUAGE GENERATION TECHNIQUES

Sign language generation techniques can be classified based on the type of input fed to the system and its corresponding output provided by the system. Sign Language generation can be classified in following two types:

- Speech/Text to Image/Video.
- Speech/Text to Animation.

A. Speech/Text to Image/Video

In this type the input to the system is either speech or text and output is an image or video of corresponding sign.

Revised Manuscript Received on July 30, 2020.

* Correspondence Author

Shruti Chikalthankar*, Computer Engineering, Pune Institute of Computer Technology, Pune, India. Email: sschikalthankar@gmail.com

Dr. Archana Ghotkar, Computer Engineering, Pune Institute of Computer Technology, Pune, India. Email: aaghotkar@pict.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Tejas and Rituparna (2016) [1] proposed a model that is capable of processing audio or text and stringing together videos and images to generate signs. The system is able to string together still images or videos for the entities not present in the repository. Input was preprocessed for removal of unwanted data such as punctuations.

N-gram algorithm was used to verify subsequence of sentence already present in repository. Parts of speech tagger was used to assign token. The dataset consists of limited words (approx 800). The words not present in repository were spelled using images and videos which make the system less efficient.

Pooja and Anita [2] proposed a model with text as input. The other functionality provided was input as an image of the sign of the alphabet or numeral. For an image as an input, the system identifies the corresponding meaning of that sign and provides output in textual format. The system uses vision based techniques to identify images. Each input image is processed to extract an array of features. These extracted features were matched with the existing set of arrays of features. If match found, the corresponding text associated with array was retrieved. The system works only for input image alphabet or numerals. Very limited database of 26 alphabet and 9 numerals were used.

Taner and Oguz [3] proposed a bidirectional system with motion capturing module to identify signs and dictate it in textual format. A voice recognition module was used for voice to sign conversion in image or video format. For speech recognition module, CMU Sphinx was used to recognize and convert speech to text. For these texts, corresponding gif images were displayed. The database was fed with 50 words of data.

Teranai and Pongpisit [4] developed a web service framework for text to gif translation. They used the longest word division method to analyze the sentence fed by the user. The output consists of a series of images for the given input. They tested system with 42,121 words with 30 sign language images. Drawback of the system observed was security of data and integrity of the system which depends on source images.

Stephanie Stoll et al uses Neural Machine Translation network based on RNN (Recurrent Neural Network) to obtain a sequence of gloss probabilities which generates human pose sequences. [5]

B. Speech/Text to Animation/Avatar

As the name suggests, the system takes speech or text as an input and displays animation as an output.

Concept of Avatar: With the development in the field of virtual reality and animation, it is now possible to create a human avatar to perform responsive signs. There are some projects that were developed for translation of English to American Sign Language. They are:

ViSiCAST (Virtual Signing: Capture, Animation, Storage and Transmission) [6]:

It was a project under the Information Society Technologies (IST) using Virtual Human technology for animation of sign language. CMU Link Parser was used to analyze input text and then prolog declarative clause grammar rules to convert this linkage output into a Discourse Representation Structure (DRS). A script of symbolic notations called as Signing Gesture Markup Language was developed which describes movement to perform sign.

TEAM Project [6]:

It was an English-to-ASL system that uses Synchronous Tree Adjoining Grammar rules to build an ASL syntactic structure while an English dependency tree was built during analysis. An ASL gloss is obtained from the linguistic portion with parameters having information of morphological variations, facial expressions, and sentence mood.

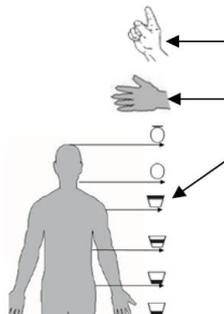
ZARDOZ Translation System [7]:

The system was developed as a cross-modal translator for English text to American, Irish and British Sign language. The system used morphological rules followed by idiomatic reduction and then parsing to produce syntactic and semantic representation. The metaphoric and metonymic structures were removed by schematization. The discourse tracking agency performs anaphoric resolution, sign syntax agency, employs spatial dependency graphs, sign mapping assigns concept-to-sign and Doll Control Language (DCL) program controls an on-screen animated doll.

III. REPRESENTING SIGN LANGUAGE AT PHONETIC LEVEL

To create signs using avatar, the avatar module must be told what to do and how to generate gestures. Most avatar modules are fed with the notational script which consists phonetic information about sign language.

Thomas Hanke proposed notations called HamNoSys (Hamburg Notation System) which describes signs at phonetic level. Notations consist of non-manual and manual information such as hand shape, hand location and hand orientation. The notations are also available for single and two handed signs along with symmetry and non-symmetry of signs. Fig. 2. shows some HamNoSys symbols and their description. [8]



Symbol	Description
☞	indexfinger stretched
△	extended finger ahead
0	palm orientated left
☞	location shoulder height
☞☞	fully stretched out
↑	hand move ahead
→	hand move right

Fig. 2. HamNoSys Symbols and Description. [8]

Several work has been done to generate a system from text/speech to Indian Sign Language (ISL). Authors of paper [9], [10], [11], [12] have used the approach of text/speech to ISL using animation module. Authors of [13] proposed a system to generate HamNoSys of ISL for given input words.

IV. METHODOLOGY

Fig. 3 Shows the general architecture of Text-to-Sign generation system.

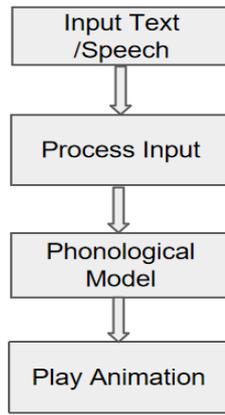


Fig. 3. Text-to-Sign language Translation

A. Processing Input

The processing techniques depend on the type of input the system allows. For the system which allows speech as input, various speech recognition algorithms can be used. A hidden Markov model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states and can be represented as Bayesian network. Dynamic time warping (DTW) is used to find an optimal alignment between two given (time-dependent) sequences under certain restrictions, the sequences are warped in a nonlinear fashion to match each other. Artificial neural network (ANN) is non-linear data driven self-adaptive approach. It can identify and learn correlated patterns between input dataset and corresponding target values.[14] If input is in textual form then techniques such as tokenization, lemmatization and parts of speech tagging must be performed.

Toqueer Ehsanand Sarmad Hussain analysed the statistical and neural parsing for Urdu. They used probabilistic context free grammar, data oriented parsing and recursive neural network based models using multiple linguistic features. Features such as syntactic sub-categorization of POS tags, empirically learned horizontal and vertical markovizations and lexical headwords were used.[15]

Zhenghua Li et al, used Coupled POS tagging method on heterogeneous annotations for Chinese POS tagging. They used multiple labeled dataset for conversion of spoken language to informal text such as tweets and product comments. The two sets of POS tags together such as for Noun NN,n were combined to build a conditional random field (CRF) based tagging model using ambiguous labeling.[16]

In order to take input from deaf and dumb user, it is also important to help user to predict the words or sentence that user wants so as make system convenient and easy to use. Akshay Bhatia et al, developed a system that can predict and correct text input for desktop editor. The predictive algorithms used were n-grams and suffix trees. Authors used “trie” data structure in which nodes stores alphabets and is traversed with respect to the frequency to suggest top words. They used n-gram algorithm for prediction of words with default frequency in phase 1 and modified frequency in phase2.[17]

Meishan Zhang et al developed a neural model for Chinese word segmentation and POS tagging. A bi-directional LSTM was used to predict next character in the sequence. Input features used was character bi-grams. The method outperforms the existing systems. [20]

Sometimes the user may not be able to enter the exact word that is present in the database. Using a keyword matching module the system will select the word from the database having the context similar to the word entered by the user.

Dunlu Peng et al, developed a system with semantic crossover for matching sentences. The model extracts the matching information of two sentences from the semantic interaction information generated from different angles and calculates the matching degree of the two sentences. [18]

B. Phonological Model

As discussed in the earlier section, to create an avatar it is required to create an intermediate notational script which will describe the gesture to be performed by the avatar. The input data needs to be mapped with corresponding notational script. To create such types of notations, the user admin must have knowledge of the corresponding sign language. User can gain this knowledge from active signers or from a dataset of video. Table 1. shows publicly available dataset:

Table – I: Publicly available sign language video dataset [19]

Dataset	Country	Language Level	Classes	Videos	Signers
RWTH Phoenix	Germany	Sentence	1200	45760	9
Boston ASL LVD	USA	Word	3300+	9800	6
DEVISI GN-D	China	Word	500	600	8
IITA-R OBITA	India	Word	23	-	-
SIGNUM	Germany	Sentence	450	33210	25
Purdue ASL	USA	Word/Sentence	-	-	5

C. Play Animation

Once notational script is identified, it can be passed to avatar to play signs. The accuracy of signs completely depends on this notational script.

V. APPLICATIONS

Sign language can be used in various application systems such as:

1. News Channel
2. Banks
3. Railway Platforms
4. Schools/Colleges
5. Hospitals
6. Hotels
7. Airports
8. Entertainment Programs



VI. CONCLUSION

This paper provides the study on sign language generation approaches and techniques. From the survey, it is observed that very limited work has been done for the hearing impaired community, especially in India. A sign generation system can be developed which can generate signs for the given input. This system can bridge the gap between hearing and speech impaired people and the normal people. With the usage of virtual reality and animation and natural language processing a dynamic system can be developed that can be made available in various sectors such as in banks, railways platforms, hospitals, schools, etc. This system can overcome the limitation of less number of human interpreters.

REFERENCES

1. T. Dharamsi, R. Jawahar, K. Mahesh and G. Srinivasa, "Stringing Subtitles in Sign Language," *IEEE Eighth International Conference on Technology for Education (T4E)*, Mumbai, 2016, pp. 228-231.
2. Pooja Balu Sonawane, Anita Nikalje, "Text to Sign Language Conversion by Using Python and Database of Images and Videos", *International Journal of Engineering Research in Electronics and Communication Engineering*, vol 5, February 2018
3. Taner Arsan and Oğuz Ülgen, "Sign Language Converter", *International Journal of Computer Science & Engineering Survey (IJCSSES)*, vol.6, No.4, August 2015.
4. Teranai Vichyaloetsiri and Pongpisit Wuttidittachotti, "Web Service Framework to Translate Text into Sign Language", *IEEE*, 2017.
5. Stephanie Stoll, Necati Cihan Camgoz, Simon Hadfield, Richard Bowden, "Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks", *International Journal of Computer Vision*, December 2019.
6. Matthew P. Huenerfauth, "A Survey and Critique of American Sign Language Natural Language Generation and Machine Translation Systems" (Technical Report MS-CIS-03-32). Philadelphia: Department of Computer and Information Science, University of Pennsylvania. September 2003.
7. Tony Veale and Alan Conway, "Cross Modal Comprehension in ZARDOZ An English to Sign-Language Translation System" , 7th International Generation Workshop, June, 1994.
8. Hanke, T., "HamNoSys - representing sign language data in language resources and language processing contexts." In: Streiter, Oliver, Vettori, Chiara (eds): *LREC 2004, Workshop proceedings: Representation and processing of sign languages*. Paris: ELRA, 2004, - pp. 1-6.
9. Tirthankar Dasgupta, Sambit Shukla, Sandeep Kumar, Synny Diwakar, Anupam Basu, "A Multilingual Multimedia Indian Sign Language Dictionary Tool ", *The 6th Workshop on Asian Language Resources*, 2008.
10. Anuja.K, Suryapriya.S, Sumam Mary Idicula, "Design and Development of a Frame Based MT System for English-to-ISL", *IEEE World Congress on Nature & Biologically Inspired Computing*, 2009.
11. Malu S Nair, Nimitha A P, Sumam Mary Idicula, "Conversion of Malayalam Text to Indian SignLanguage Using Synthetic Animation", *International Conference on Next Generation Intelligent Systems (ICNGI)*, 2016.
12. Purushottam Kar, Madhusudan Reddy, Amitabha Mukerjee and Achla M. Raina, "INGIT: Limited Domain Formulaic Translation from Hindi Strings to Indian Sign Language", *International Conference on Natural Language Processing (ICON)*, 2007.
13. Khushdeep Kaur and Parteek Kumar, "HamNoSys to SiGML Conversion System for Sign LanguageAutomation", *Twelfth International Multi-Conference on Information Processing*, 2016.
14. Pahini A. Trivedi, "Introduction to Various Algorithms of Speech Recognition: Hidden Markov Model, Dynamic Time Warping and Artificial Neural Networks", *International Journal of Engineering Development and Research*, vol 2, 2014.
15. Toqueer Ehsan and Sarmad Hussain, "Analysis of Experiments on Statistical and Neural Parsing for a Morphologically Rich and Free Word Order Language Urdu", *IEEE Access*, October 2019.
16. Zhenghua Li, Jiayuan Chao, Min Zhang, Wenliang Chen, Meishan Zhang, and Guohong Fu, "Coupled POS Tagging on Heterogeneous

Annotations", *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, March 2017.

17. Akshay Bhatia, Amit Bharadia, Kunal Sawant, Swapnali Kurhade, "Predictive and Corrective Text Input for desktop editor using n-grams and suffix trees", *International Conference on Advances in Human Machine Interaction*, 2016.
18. Dunlu Peng, Shaohong Wu, Cong Liu, "MPSC: A Multiple-Perspective Semantics-Crossover Model for Matching Sentences", *IEEE Access*, May 2019.
19. "Sign language datasets", Available: https://facundoq.github.io/unlp/sign_language_datasets/index.html
20. Meishan Zhang, Nan Yu, Guohong Fu, "Simple and Effective Neural Model for Joint Word Segmentation and POS Tagging", *IEEE/ACM Transactions on Audio, Speech and Language Processing*, Vol 26, September 2018.
21. Jane E. Johnson and Russell J. Johnson, "Assessment of Regional Language Varieties in Indian Sign Language", *SIL Electronic Survey Report*, April 2008.

AUTHORS PROFILE



Shruti Chikalthankar currently pursuing Masters Degree program in Computer Engineering, in Pune Institute of Computer Technology, Pune (M.S), India. Her research interest includes Natural Language Processing, Image Processing and Human Computer Interaction.



Dr. Archana Ghotkar has received B.E degree in Computer Science and Engineering in 1998 from Dr. BAMU university, M.E degree in Computer Engineering in 2000 and Ph.D in Computer Engineering from Savitribai Phule Pune university. She is an associate professor in the department of computer engineering at Pune Institute of Computer Technology in Savitribai Phule Pune university, INDIA. Her main research interest includes Computer vision, Pattern Recognition, Human computer interaction and Data mining. She is a member of ISTE and CSI.