Multimedia Tools and Applications (2022) 81:4283–4321
https://doi.org/10.1007/s11042-021-11706-1

An automatic machine translation system for multi-lingual speech to Indian sign language

Amandeep Singh Dhanjal[1] · **Williamjeet Singh**[1]

Received: 28 May 2021 / Revised: 25 August 2021 / Accepted: 22 October 2021 /
Published online: 6 December 2021

Abstract

Sign language (SL) is the best suited communication medium for hearing impaired people. Even with the advancement of technology, there is a communication gap between the hearing impaired and hearing people. The aim of this research work is to bridge this gap by developing an automatic system that translates the speech to Indian Sign Language using Avatar (SISLA). The whole system works in three phases: (i) The first phase includes the speech recognition (SR) of isolated words for English, Hindi and Punjabi in speaker independent environment (ii) The second phase translates the source language into Indian Sign Language (ISL) (iii) HamNoSys based 3D avatar represents the ISL gestures. The four major implementation modules for SISLA include: requirement analysis, data collection, technical development and evaluation. The multi-lingual feature makes the system more efficient. The training and testing speech sample files for English (12,660, 4218), Hindi (12,610, 4211) and Punjabi (12,600, 4193) have been used to train and test the SR models. Empirical results of automatic machine translation show that the proposed trained models have achieved the minimum accuracy of 91%, 89% and 89% for English, Punjabi and Hindi respectively. Sign language experts have also been used to evaluate the sign error rate through feedback. Future directions to enhance the proposed system using non-manual SL features along with the sentence level translation has been suggested. Usability testing based on survey results confirm that the proposed SISLA system is suitable for education as well as communication purpose for hearing impaired people.

Keywords Indian sign language · Speech recognition · Machine translation · Multi-lingual speech · HamNoSys · Speech to Indian sign language

✉ Amandeep Singh Dhanjal
aman.dhanjal13@live.com

Williamjeet Singh
williamjeet@gmail.com

[1] Department of Computer Science, Punjabi University, Patiala, India

1 Introduction

Communication is considered as one of the ways for human beings to convey messages, express their feelings, ideas and information with other persons. Verbal/Vocal communication and Non-verbal/Non-vocal communication are the two categories of communication mode. Verbal communication is a type of oral communication wherein the message is conveyed through spoken words and thus known as speech communication. Non-verbal communication conveys the message without written or spoken words. Especially, hand movements are used to perform a gesture in non-verbal communication. Speech is understood through the power of hearing but due to various types of accidents and injuries, there is an increase in the number of physically disabled people at the early age of birth. The hearing power loss occurs due to the lack of perceiving the sound elements such as pitch, loudness, timbre and frequency of the sound [23]. Due to hearing impairment some people are unable to listen to speech (completely or partially), so they use sign language for communication. Juan Pablo de Bonet published the first manual alphabet book in 1620 that helps to learn sign language [49, 114]. But hearing people are not familiar with sign language and they are unable to understand the sign gestures. There are around 500 schools available in India for hearing impaired people, which are less in number as compared to the hearing impaired population [90].

Sign language interpreter plays a significant role to bridge the gap between hearing impaired and hearing people. Bi-directional communication in the machine translation, system comprises of two sub-systems: (i) receives the speech or text as input and generates the corresponding gesture as output, (ii) captures the gesture as input and generates the text or speech as output [8, 84]. As per Ethnologue 2019, there are 7111 spoken languages and near about 200 sign languages in the world [3, 45]. Approximately 6 million hearing impaired people are living in India [90]. This is also one of the assumptions that if the communication barriers are removed for the hearing impaired in the education system, they should learn at the same rate as hearing students learn [68].

1.1 Automatic speech recognition

Research on Automatic speech recognition (ASR) has drawn a lot of attention over the last five decades and therefore the development of its applications is quite optimistic [87, 92]. ASR is the process that analyzes the audio signals (produced through microphone or audio file), extracts the sound features and then generates the text as a result. Every spoken word is divided into segments and every segment is further composed of several formant frequencies [103]. A formant is a spectral shape that results from an acoustic resonance of the human vocal tract. Every formant has its bandwidth and amplitude. Feature extraction techniques such as MFCC, LPC, PLP are used to extract feature vector from the input sound. Speech has featured in both the time domain and frequency domain. In time-domain various features are energy of the signal, short-time zero-crossing rate, maximum amplitude, minimum energy and autocorrelation. The frequency-domain feature is Short-time Fourier transform, Wideband spectrum, Narrowband spectrum [22, 66]. Observation is stored in acoustic vector and decoded using Hidden Markov Model (HMM). The likelihood is decided by an acoustic model and a language model. The acoustic model supplies the possible phonetic sequence for a word and language model to make a sentence based on the n-gram model.

1.1.1 Classification of ASR systems

ASR systems are classified based on the types of speakers, utterances and vocabulary size. According to speaker variability, ASR system is divided into three types: speaker dependent, speaker adaptive and speaker-independent [53, 117]. Speaker dependent speech recognition systems work well only for those users whose audio recordings are used earlier to train the system. As per the named speaker adaptive, these types of systems have the features to adopt new users even if the model is trained with limited data and limited users. Google speaks, Siri, Microsoft Cortana are the important examples of the present day's automatic speech recognition system.

These systems are trained with a high volume of data and require remarkably high configuration resources to train the system. Speaker independent automatic speech recognition systems work for every user whether their voice is used during the training process or not. Based on vocabulary size ASR systems are divided into four categories:

- *Small:* corpus contains 1-100 unique words
- *Medium:* corpus contain 101-1000 unique words
- *Large:* corpus contains 1001-10,000 unique words
- *Very Large:* more than 10,000 unique words

Based on pronunciation, ASR systems are classified into four categories: isolated, connected, continuous and spontaneous. It is easy to extract the features for isolated words because these words are pronounced with a silence both at the start and end of the speech. When words are joined and pronounced with little pause/silence, they are categorized as connected words. It is a natural form of human speaking and it is difficult to recognize properly. In this type of speech, human speaking continues like a complete sentence. Spontaneous speech is optimal for human-human communication but it is difficult for machine translation. Spontaneous speech includes duplications, faltering and incomplete vocabulary.

1.1.2 Speech recognition models development tools

There are various tools available to train the new speech recognition models rather than building a new tool from scratch. The training process requires raw data as prescribed by the given tool. The most preferable open-source speech recognition tools and their supported programming languages are listed below:

- Kaldi [110]: C++, Python
- DeepSpeech: Python, JavaScript, Go, Java,. NET
- CMU Sphinx [1, 40, 93, 100]: Python, C, Java
- HTK [75, 97, 99]: C, Python

1.1.3 Challenges for the development of speech recognition model

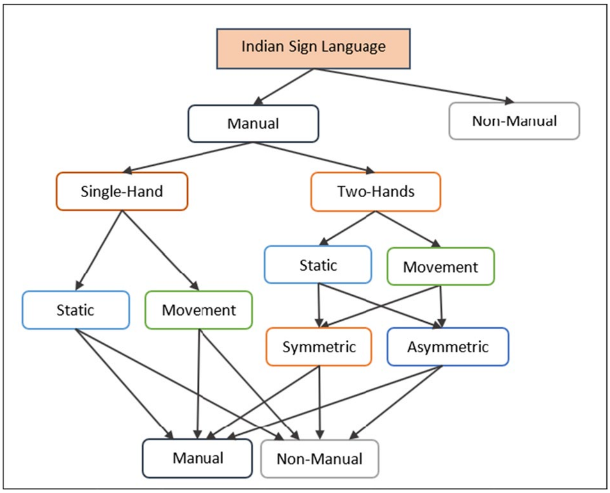
- *Contextual difference*: homophone words such as “Son” and “Sun”, “write” and “Right” are almost the same in English (Indian) pronunciation but are different from their meaning [17, 58].
- *Style variability:* fluency of speaking style affects the information available in both time domain and frequency domain of speech signal [57, 67].
- *Language Model*: In machine learning, speech recognition system works only for the spoken language that is used to build the model
- *Speaker dependency:* Speaker independent ASR models are difficult to build because number of variations exist in speech signals due to diversity of age group, gender, pitch and tone. [57, 109].
- *Background noise:* Inclusion of extra audio signals such as construction activity, Barking dog, music, noisy conflict nearby and horns also affects the system and degrades the accuracy level of speech recognition [28].

1.2 Sign language

Sign language or gesture language is used to share information between source and sender in the same way the other spoken languages are used. It is a complete visual mode of communication and understood through the vision power rather than the hearing power. Each sign language has the following phonological features: handshapes, location of hand, movements body parts (specially hands) and facial expressions.

Sign language is the third most used language in the United States and the fourth most used language worldwide. Indian sign language (ISL) is the native language of the Indian Deaf community. Hearing impaired people use ISL as the primary form of communication in their daily lives. Sign Language has its own syntax and grammar structure to perform a gesture corresponding to spoken word [23]. Figure 1 displays the hierarchy of ISL types based on various parameters.

Fig. 1 Hierarchical division of Indian Sign Language



- **Manual Sign:** Hand(s) shape, motion and location parameters are included in manual signs. These signs are performed with a single hand or with both hands but do not include extra information about emotions. ISL numbers, fingerspell words and alphabets are the common examples of manual signs.
- **Non-Manual:** Non-manual gestures include shoulder-shrugging, bending body, movement of body, eyes, eyebrows, cheeks and mouth features with or without hand gestures. For example (*happy*, खुश, ਖ਼ੁਸ਼), (bed, ਪਲੰਗ, बिस्तर) words include facial expression along with manual sign in ISL.
- **Single-hand Sign:** Includes only a single dominant hand to perform a gesture of sign alphabet or word. For example, a sign of numeric numbers from 0 to 9 using only a single hand and after 9th number, the signer has the choice to use the same dominant hand or both hands for two or more digits. The dominant hand is the most commonly used hand in daily life activities such as eating, writing as well as doing other tasks. It is used to perform single-hand signs, two-hand symmetric signs and two-hand asymmetric signs.
- **Two-hand Sign:** Sometimes the single hand is not sufficient to explain the properties of a particular word. In that case, both the hands are used to perform the gestures. For example, word (bottle, ਬੋਤਲ, बोतल), (hut, ਝੌਪੜੀ, झौंपड़ी), (baby, ਬੱਚਾ, बच्चा).
- **Tow-handed symmetric signs** are performed with both the hands either simultaneously or alternative but the handshape, movement and location must be the same for both.
- **Two-handed asymmetric signs** are performed with the help of both hands where dominant hand (right) performs the key role of gesture and passive hand (left) serves as base or helper.

1.2.1 Various notation systems

A notation or script is a set of symbols, shapes or characters used to make a writing system visible. On the other hand, sign language does not have any standard written form. An earlier method of writing a sign language has been developed by William Stoke in 1960 specifically for American sign language. Stokoe is a phonetic notation system based upon three major parameters: (i) hand location, (ii) handshape and (iii) hand movement [33, 46, 61]. Valerie Sutton has developed the SignWriting notation system which includes both manual and non-manual features of a sign gesture. SignWriting notation system is a group of more than 630 symbols which are collectively written in the pictorial form [6, 13, 71]. An incredibly useful tool JSPad has been developed to write and generate SignWriting signs for the Japanese Sign language [71]. HamNoSys notation is most widely used for research purposes and animate gestures with a 3D avatar and recently used to develop text to Sign language translation [25, 41, 108]. HamNoSys stands for the Hamburg Notation System which was developed in 1984 by Thomas Hanke at the University of Hamburg [33, 46, 47]. The first version has been updated many times with the inclusion and exclusion of various symbols in the notation system. Currently, HamNoSys - 4.1 is available that supports manual and non-manual sign notations having around 230 symbols. Gloss is another notation system that is purely written in simple words [7, 69, 114].

2 Objectives of research work

So far, most of the work has been done for text to sign language but less work has been concentrated on multi-lingual speech to the ISL translation system. It is the first automatic machine translation system that supports multi-lingual speech as input which is further converted into ISL. However, speech recognition models are designed only for spoken isolated words in English, Punjabi and Hindi language. The focused area of the proposed research is to build an automatic machine translation system for the conversion of speech to Indian Sign Language using Avatar (SISLA). The objectives of research work are listed below:

(i) To develop a Hidden Markov Model (HMM) based speech recognition models for three different languages (English, Punjabi and Hindi).
(ii) To generate an ISL corpus for at least 400 words in each language using HamNoSys notation system.
(iii) To convert a recognized word into a sign language notation system using markup language (SiGML) that provides effective machine translation of sign gestures over the web platform.
(iv) To represent an ISL gesture animation in 3D view space that provide 360^0 view rotation of signing avatar.

3 Related work

Nowadays, speech is not only limited to text conversation [40, 79, 99] but also applicable in a variety of machine interactions such as security purposes (device access restriction [79, 87], alert systems [20, 59, 94], content search [50], dynamic video caption [34, 35]) and translation systems (speech to text, speech to another language).

An interactive speech to sign language translation system named as TESSA (Text and Sign Support Assistant) has combined speech recognition technology and virtual avatar animation for the communication system in the UK Post office. It has used two approaches: The first one is phrase lookup and the other one is Sign-Supported English (SSE) to translated clerk's speech into British sign language [21, 115]. SASL-MT (South African Sign Language Machine Translation) project has been developed to bridge the gap between hearing and hearing impaired people living in South Africa [112]. This system has primarily focused on the generation of gestures using non-manual features. INGIT (a Sanskrit word) is a small domain corpus that has been implemented to translate Hindi text to ISL. Gloss string is generated after syntactic and semantical analysis with Fluid Construction Grammar (FGC) [43]. ATLASLang machine translation system makes use of signing avatar to translate Arabic text to Arabic sign language. A morpho-syntactic analysis is performed based on rule-based and example-based approaches for the input sentence and displays the sequence of sign language gif images [16]. A virtual keyboard is designed to translate an Arabic sentence into Arabic sign language in the form of sign images [4].

Audio-Visual media content is easily shared through Digital television (TV) and its useful for all types of users [95]. A TVML (Television program Markup Language) script has been used to develop a new 3D avatar for the Television program. Optical motion capture technology is used to generate the 4900 sign word [42]. A quite different wearable device VisAural is developed for hearing impaired-people which show the direction of sound source [30]. Web-based e-learning system has been developed in the context of education for hearing impaired people. The system focuses on the research to translate English text into Pakistani sign language [24]. Spanish speech to Spanish sign language translation system focuses on the sentences related to the purpose of applying or renewing the identity cards in the office. System decodes spoke utterance into a sequence of words which are further processed by a natural language translator. To generate a Spanish sign gesture, 3D avatar (VGuido) is used which uses SiGML script file as the sequence of gesture parameters. Automatic speech translation system has been trained using Hidden Markov Model with 416 sentences and more than 650 words. From a voice sample, 13 perceptual linear prediction coefficients are derived from a Mel-scale filter bank. The analysis has been carried out as per semantic concepts of words and signs rather than a translation of words and signs directly. In restricted domains, the rule-based strategy provides better results and statistical models need a high volume of data for the training process [96]. A novel approach for assisting hearing impaired people effectively has been proposed where a bidirectional dialogue enabled virtual agent system has been developed. The system supports the functionality of both speech-to-text and text-to-speech [60] conversions.

Suvarna et al. has proposed a rule-based machine translation system for Marathi language to ISL translation that support both text and speech input [10]. Nayan et al. has developed a machine translation system for Indian Sign Language (ISL) using Gallaudet University's database [73]. The proposed system has been divided into two sub-sections: (i) display video captions in English language, (ii) second section plays a 3D avatar video for English words. Navroz and William have presented a review study about the machine translation of text to sign language and highlights the benefits and importance of the 3D synthesis sign gestures [39]. A hybrid approach for synthesis sign has been proposed to create a synthetic sign gesture using 3D avatar. In this technique motion capture data of French Sign Language is used to generate new synthetic signs [81]. Research work offers the dual-mode of communication using Microsoft Kinect v2 device for hearing impaired people. Sign language to speech and speech to sign language with a 3D animated avatar along with the subtitles have been proposed [2]. State-of-the-art speech technology for live subtitle system has been designed for public service to be broadcasted in the Czech Republic Television [89]. Bangla speech to sign language translator has been developed using CMU Sphinx tool and macOS. The spoken word is broken down into chunks of frames which are phonetically identified based upon their feature vector. The acoustic and language models decode the words available in the trained model's dictionary. In the last step, images of sign language are displayed corresponding to the recognized Bangla text [100]. A context-based machine translation system for Tamil speech to ISL has been implemented using the CMU sphinx toolkit. Priyanka and Arvind have used an avatar approach to display the sign gesture corresponding to a spoken phrase [90]. A real-time speech to sign language prototype has been proposed over the webserver to facilitate the Hearing impaired community through smartphone video streaming. CMU PocketSphinx 4 toolkit is used to recognize English spoken words to English text in the mobile device. To generate a language model, the SRI Language Model (SRILM) tool is used. Phonetically Tied Mixtures (PTM) having 5000 senones and 128 mixtures acoustic model is used to provide better decoding speed and accuracy in low processing power devices [86]. Real time note taking assistance in the classroom for hearing impaired people has been proposed using Microsoft speech recognition utility application. Valanarasu has used the dictionary building software tool to add a custom dictionary in speech recognition [52].

At present, the number of social and networking sites such as Twitter, Facebook and Instagram have grown rapidly for various communication purposes [111]. Various authors performed text analysis on the test data collected from these websites for the research purpose and presented different predictions such as personality predictions, happiness, well-being, emotion mining and opinion mining [102, 107]. But in sign language there is less content available for the research and thus this domain needs much more attention. ViSiCAST text to 3D animation of sign language system performs translation from spoken language to three distinct national sign languages: British Sign Language, Dutch Sign Language and German Sign Language [26]. SignDict is an OCR based text to ISL application that has been built for the mobile operating system. Video-based sign language dataset developed by Indian Sign Language Research and Training Center (ISLRTC) has been used in this application [38].

Limitation of Nayan et al.'s work is analyzed for text to sign language as the system generates the static animated video signs. This is a time consuming process to modify the sign language parameter after exporting the sign video [73]. The research work of Lucie et al. is limited to the phonological components (hand shape, hand location and hand movement) [81]. After reviewing the past systems, using CMU Sphinx for creating the language model has been proven to be the most efficient approach.

4 Contribution

SISLA is a web-based language translator which aims at the translation of isolated spoken words into Indian Sign Language (ISL). In order to achieve the objectives of this study, the key contributions of this research work are listed below:

1. Created a speech corpus for English, Hindi and Punjabi spoken language.
2. New models have been trained for speech to recognize spoken isolated word in all the three languages.
3. HamNoSys notation system has been used to generate synthetic corpus for ISL.
4. Implemented a system to automatically translate speech into ISL using 3D avatar.
5. Comparative analysis of developed system with other existing machine translation systems for sign language.

5 Proposed system

An attempt has been made in this research work to develop a multi-lingual speech recognition system that converts a speech into the corresponding ISL. The flowchart of the proposed system is displayed in Fig. 2.

The speech processing task is divided into two functional stages: feature extraction and the decoding stage.

(i) **Feature extraction**: The different steps involved in the pre-processing stage are as follows: *voice input*, where data is processed through a microphone, *Pre-emphasis,* where the speech signals are improved, *Frame Blocking,* where the speech signals are divided into several subunits and *Windowing,* where the signal discontinuities are minimized.

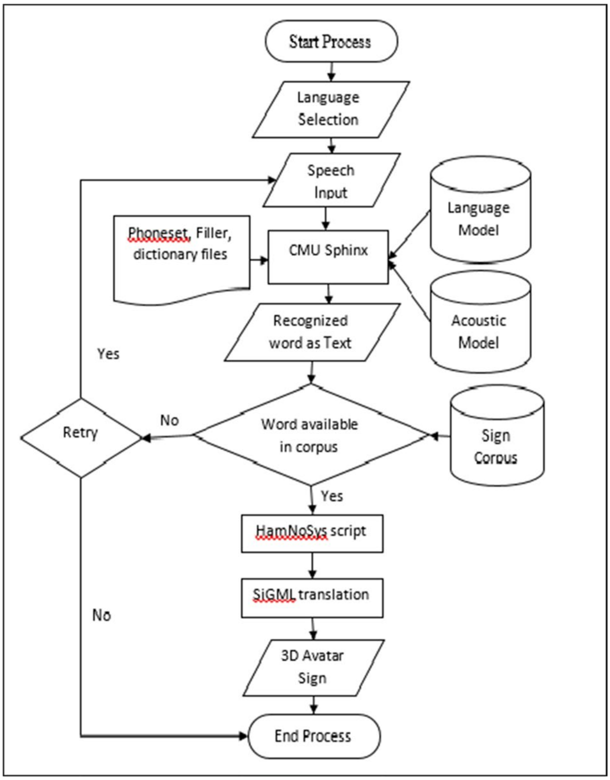


Fig. 2 Flowchart of proposed SISLA

> A fundamental problem of speech processing is to characterize audio waves in terms of speech signal models.

Speech signal models provide us the theoretical description of audio signal processing to produce the desired output. Pre-emphasis of audio signal processing boosts the amount of energy in the high frequencies; Thus, the recognizer gets more formants for the acoustic model. Voiced signals such as vowels (English: a, e, i, o, u; Hindi: अ, आ, इ, ई, उ, ऊ, ऋ, ए, ऐ, ओ, औ; Punjabi: ਅ, ਆ, ਇ, ਈ, ਉ, ਊ, ਏ, ਐ, ਓ, ਔ) have higher amplitude than consonants. Also, the noise has higher energy as compared to the vocal sounds. In the pre-emphasis phase, the noise reduction is done with the help of following equation:

$$\begin{gathered} y(n) = x(n) - \alpha x(n-1) \\ 0.9 < \alpha < 1 \end{gathered} \tag{1}$$

In the above equation $y(n)$ is the output function and the typical value of filter constant α is between $0.9 < \alpha < 1$ which takes the derivative in the time domain from n to n-1 sample.

Another reason to use pre-emphasis is that the numerical values of the low-frequency signal tend to change slowly or seamlessly from sample to sample. Thus, the high-frequency signal changes rapidly which provide the best comparison between two samples.

In the next step each $x(n)$ speech signal is divided into 20 ~ 30 ms frames of N samples and the adjacent frames are separated by M samples where M < N. Typical values of M = 100 and N = 256 are chosen to get sufficient information from small frames. The first frame consists of first N samples, the second frame overlaps N - M samples, the third frame overlaps N - 2 M samples and continue until the input speech is processed completely. To minimize the signal discontinuity at the beginning and end of each frame, the windowing step is performed. Hamming window is used as depicted in the following equation:

$$y(m) = x(m) * w_n(m) \tag{2}$$

In the above equation $y(m)$ is the output signal function, $x(m)$ is the input signal and $w_n(m)$ is the window function where m is $0 \le m \le N_m - 1$ and N_m stands for the number of samples within each frame. The window function is defined as:

$$\begin{gathered} w(m) = \alpha - (1 - \alpha)\cos\frac{2\pi m}{N-1} \\ 0 < \alpha < 1 \\ \alpha = 0.54 \, \textit{for Hamming window} \end{gathered} \tag{3}$$

Computation of Discrete Fourier Transformation (DFT) is done using a Fast Fourier Transformation (FFT) function that converts the power spectrum of window data into a frequency spectrum. Following equation for Fourier Transform is followed on the given set of N_m sample defined as:

$$x_n = \sum_{m=0}^{N_m - 1} x_n e^{-j2\frac{\pi km}{N_m}} \tag{4}$$

Warp the DFT output to the Mel-scale: Mel is a unit of pitch such that the sounds which are perceptually equidistant in pitch are separated by the same number of mels. Mel filter bank is computed from the following equation.

$$F_{mel} = 2595\log\left(1 + \frac{F_{HZ}}{700}\right) \tag{5}$$

(ii) **Decoding process:** In the decoding phase, computation is performed to find the sequence of words which is a most probable match to the feature vectors. HMM is defined as a tool comprising of a finite set of states and transition probabilities at a regular time interval. Every state in HMM has a probability outcome which is distributed between 0.0 to 1.0. The probabilistic pattern is defined to match the word which is based on probability sequence of observations in HMM as:

$$O = O_1, O_2, O_3, O_4, \ldots \ldots \ldots O_n \tag{6}$$

These are the observations obtained from the feature vectors generated by a feature extraction process using MFCC as already discussed in the previous section. Speech recognition is predicted in HMM using the following equation:

$$P(W|O) = \frac{P(O|W) * P(W)}{P(O)} \tag{7}$$

For this step, three important components must be present; an acoustic model for each unit (phoneme or word), dictionary file containing words and their phoneme sequences and a language model with words or word sequences likelihoods. Figure 3 illustrates the above-mentioned equation and basic decoding architecture of HMM.

During the training process, it is the most challenging task for the HMM to fine-tune the model parameter that maximizes the likelihood of the observation sequence. There is no optimal way for this problem but the **re-estimation** of the HMM parameter helps to optimize the probability.

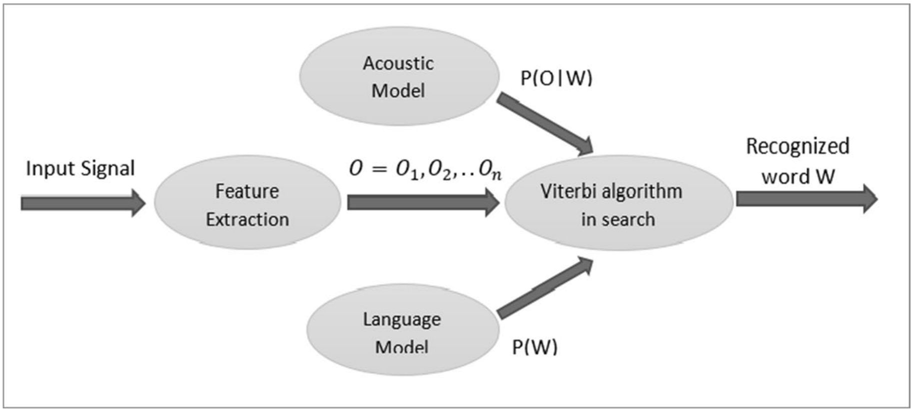


Fig. 3 Basic decoding architecture of HMM

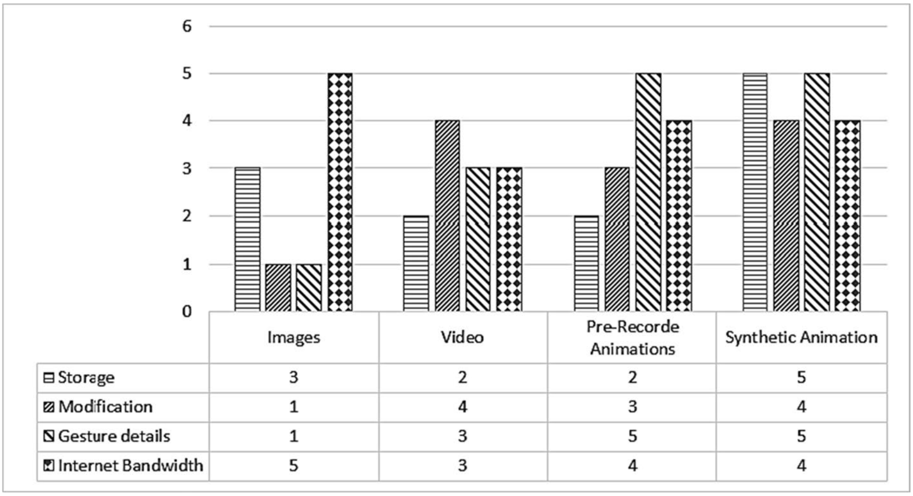
In the Forward algorithm approach, input is divided into sub-sequence of states, then the probabilities are computed and finally store them in a table for later use. The probability of a larger sequence is obtained by combining the probabilities of these smaller sub-sequences. There are several paths through the hidden states in the HMM that provide the possible sequences. Once the speech recognition model is trained successfully, a trained model is able to connect with the developed system.

5.1 Sign language processing

Images do not provide sufficient information about sign gestures thus hearing impaired people prefer to use sign video representation rather than sign images. While the contents are being shared over the web, the videos need a high speed internet connection and larger storage space [88]. Image capturing cameras and editing software are required for capturing and modifying the sign gesture image. The same video recorder and distinctive feature software are used to create or edit sign gesture video data. Pre-recorded 3D-animations require special software (Maya, Blender, 3D-Max) that supports the keyframe editing of animation components (bone joints) [9, 19, 42, 54, 74, 104]. Sometimes motion capturing hardware is required to generate animations [63].

Synthetic animations require specific scripting font and tool that render the 3D avatar. Sign language script notation with a 3D synthetic animation scheme is selected for SISLA which is a satisfactory solution for hearing impaired people as suggested in previous research [18, 37, 44, 51]. Figure 4 displays the comparison of various approaches for sign language representation through web service. The ranking between 1 to 5 has been used to generate the results whereas:

- For storage: 1 represent high volume and 5 as low
- Modification: 1 represents the extremely low possibility and 5 as easily acceptable changes
- Gesture details: 1 represents the low information and 5 clearly explains in 3D view space
- Internet Bandwidth: 1 shows the high-speed requirement and 5 as a low transfer rate



	Images	Video	Pre-Recorde Animations	Synthetic Animation
⊟ Storage	3	2	2	5
▨ Modification	1	4	3	4
⧅ Gesture details	1	3	5	5
⊠ Internet Bandwidth	5	3	4	4

Fig. 4 Comparison of sign language gesture representation schemes

Table 1 Comparison of commonly used sign language notations

Feature parameters	Stokoe	SignWriting	HamNoSys
Year of development	1965	1974	1985
No of Symbols	55	639	230+
Manual Sign	Yes	Yes	Yes
Non-manual sign	No	Yes	Yes
Way of writing	Linear Left to right	Pictorial	Linear Left to right
Font Name	Stokoe tempo font	ISWA font	HamNoSys Unicode font
Level of Difficulty	Difficult	Easy	Medium

As per survey, the results show that the synthetic animation approach is best suited for large corpus of Indian sign language. One more concern for selection of synthetic animation approach is easy editing and correction of gesture parameters. For synthetic animation approach, an intermediate writeable notation system is required. Table 1 shows the comparison of most commonly used notation systems. After an analysis, it has been observed that HamNoSys is best suitable to represent gesture animation synthetically in association with SiGML markup language.

6 Methodology

This section explains the methodology used for building the proposed system and discuss the experimental results. Figure 5 displays the modules and sub-modules of or methodology.

The methodology is divided into four major modules: requirement analysis, data collection, technological development and evaluation.

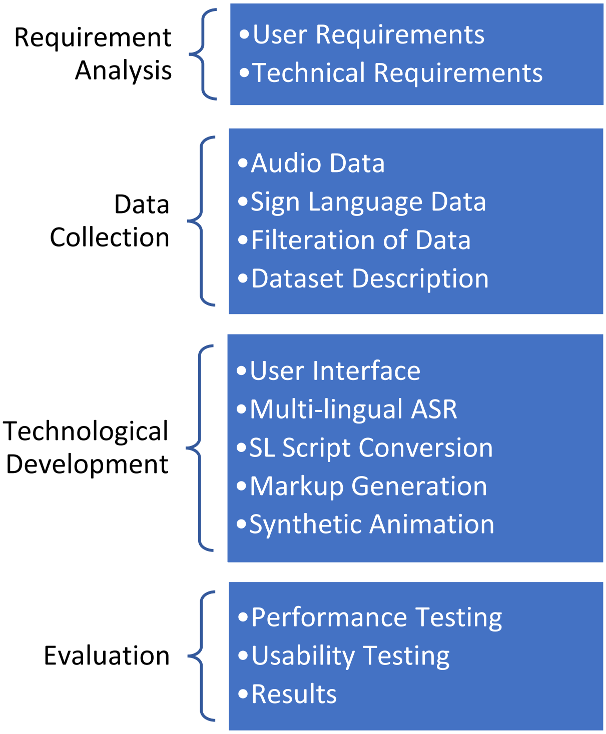


Fig. 5 Methodology used for the development of SISLA

6.1 Requirement analysis

This section describes the process of analyzing the major needs of hearing impaired users concerned with the development of SISLA machine translator system. SISLA helps to bridge the communication gap between hearing impaired and hearing people. The section is further divided into two sub-sections: user requirements and technical requirements.

6.1.1 User requirements

According to the World Health Organization, around 360 million people are hearing impaired in the world [5]. Hearing impaired and hearing people have a major gap between their source of communication [24, 74]. A major reason for the communication barriers is that majority of hearing impaired users are illiterate [11, 27, 69, 80, 85]. To overcome the communication gap between the hearing impaired and hearing people there is an urgent need of translator(s) [62, 77]. Human translators are available only in major public areas or people hire translators as per their needs.

But sometimes human translators are unable to fill the communication gap due to geological distance, language variability, lack of ability, or absence of faith [31, 56, 64]. As per the World Federation of Deaf (WFD) survey in 2009, there are 13 countries that do not have any provision of sign language interpreters [11, 29]. The World Wide Web resources are considered to be the important tools and their demand is increasing with time [23]. But there is a lack of web resources related to learning and translation in sign language [9, 14, 24, 70, 78, 116]. To overcome these limitations machine translator is required [16]. SISLA translator is the first one to overcome the gap in the context of education and communication for hearing impaired people. All these aspects and support need to generate innovative technologies to develop automated translation systems to convert this information into sign language.

6.1.2 Technical requirements

The previous section discussed the user's requirements for the development of SISLA. In technical development, there are some factors like performance, reliability and availability that SISLA must achieve a minimum threshold value to complete the research work. The main assumption imposed by the proposed system is to include isolated words as input in all the three languages: English, Punjabi and Hindi. The following technical requirements are defined to measure the quality of the system:

- ASR system must accept speech input directly in English, Hindi and Punjabi language.
- Sign language gestures should be in a synthetic 3D animation scheme.

6.2 Data collection

This section focuses on the construction and evaluation of the speech corpus in audio format for English, Hindi and Punjabi languages along with the sign language content for synthetic animations. To develop the SISLA, a list of commonly used words has been selected for all the supported languages by considering various categories (such as numerical counting, days name, months name, fruits, colors, body parts and vegetable names). Speech data has been collected from different speakers and sign language data has been prepared using different online as well as offline resources mentioned in the next sub-section. This section is divided into four sub-sections where the first two are related to the collection of data, the third elaborates the filtration of data and the last sub-section analyzes the statistics of the final corpus.

6.2.1 Audio data

Accuracy of speech recognition system is affected by variability in speakers, gender and environmental conditions. To attain a robust trained speech recognition model, big corpus of transcribed speech data is required. In the proposed system, speech corpus has been created in English, Hindi and Punjabi language from a variety of speakers. Every unique word has been recorded from around 100 speakers. Smart recorder android phone application plays a vital role during the recording phase. This application has features like auto gain, noise reduction and wav format audio file. Commonly used words such as numerical counting, days name, months name, fruits, colors, body parts and vegetable names have been selected for audio recording.

6.2.2 Sign language data

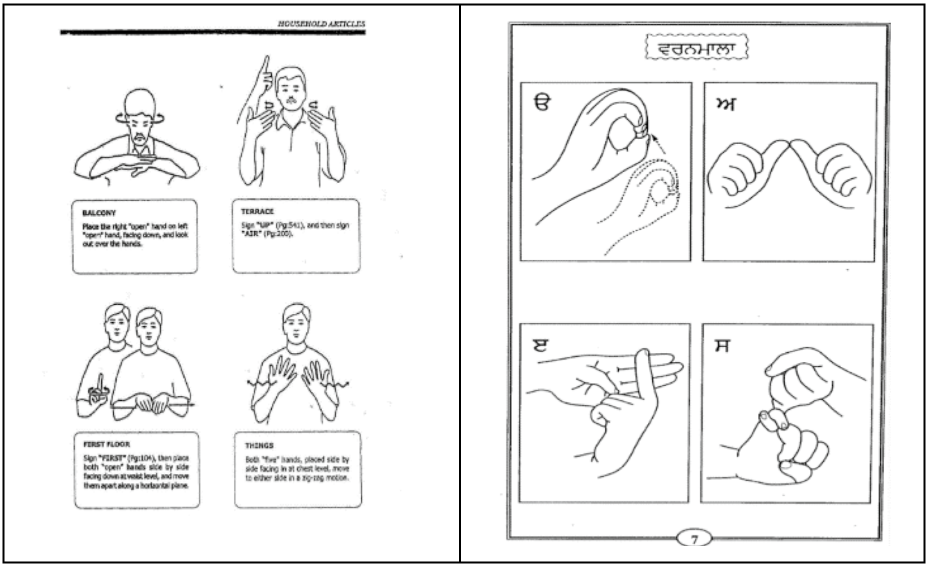
ISL gesture generation system has been developed that's includes a detailed description of sign language. There is no such standardized corpus available for ISL [106]. So, the corpus has been created with the help of experts working in Patiala School for the Deaf (Punjab, India), sign language books and from other authentic web sources. ISL corpus of daily used words (same already recorded for ASR) has been captured in HD (high definition) video format using Canon mark-II video recorder. A detailed description of the sign has been written in text format for later use in synthetic corpus. The same gesture has been recorded iteratively to include sign gestures from different view angles.

The video resources prepared by ISLRTC (Indian Sign Language Research and Training Centre), RKMVERI (Ramakrishna Mission Vivekananda Educational and Research Institute) and talking hands has been considered during the generation of ISL corpus. Study material shown in Fig. 6 has been taken from the two different Punjabi books: ਪੰਜਾਬੀ ਸੰਕੇਤਕ ਭਾਸ਼ਾ ਭਾਗ - 1, ਪੰਜਾਬੀ ਸੰਕੇਤਕ ਭਾਸ਼ਾ ਭਾਗ - 2 that has helped us to understand and develop synthetic sign gestures for ISL.

For perfect translation and synchronization of English, Punjabi and Hindi language, Oxford dictionary has been followed that provides the content translation from English to Punjabi and Hindi language along with grammar description (parts of speech tag) of a particular word.

6.2.3 Filtration of data

Speech signals are classified as voiced and unvoiced [66]. Voiced signals are produced when vocal cords vibrate during pronunciation (e.g., ‘a’, ‘b’). Unvoiced signals are non-periodic and are produced when air passes through the vocal tract during consonant spoken. In human beings, changes in size and shape of the oral cavity by the movement of articulators (jaw, tongue, lips) produce different sounds. Data filtration is required before training a speech recognition model [72]. A freeware “NCH wavepad” and “Audacity” window applications has been used to edit audio files and break down the audio file into a group of approximate 5-6 words. These softwares have features like: gain and pitch control,



HOUSEHOLD ARTICLES

BALCONY

Place the right "open" hand on left "open" hand, facing down, and look out over the hands.

TERRACE

Sign "UP" (Pg:541), and then sign "AIR" (Pg:200).

FIRST FLOOR

Sign "FIRST" (Pg:104), then place both "open" hands side by side facing down at waist level, and move them apart along a horizontal plane.

THINGS

Both "five" hands, placed side by side facing in at chest level, move to either side in a zig-zag motion.

Fig. 6 Samples of the ISL gestures from books

silence and normalization of speech signals. The average length of 15 s for each sample file has been used, which supplies optimal performance of machine during training because the longer length of audio file degrades the system performance. Each audio file is recorded in wav file format with 16,000 sample rates.

6.2.4 Dataset description

This section shows the statistics of both speech and sign language corpus that have been included in this study. To create a robust speech recognition system, various parameters such as variability in age, gender, geographical area, environmental conditions have been examined while preparing speech data. Speech data is unbiased, both male/female along with different age group speakers have been selected. After filtration and processing of data 425 (English), 413 (Hindi) and 407 (Punjabi) unique words have been selected to build ASR models for SISLA development. Table 2 shows the detailed description of speech corpus in more detail.

Each audio sample has maximum 5 words and each word is separated with average ~ 300 ms silence between two consecutive words. Total sample data is divided into 75% for training and 25% for testing. So, two separate training and testing sets have been prepared that contains English (training: 12660, testing: 4218), Hindi (training: 12610, testing: 4211) and Punjabi (training: 12600, testing: 4193) sample files.

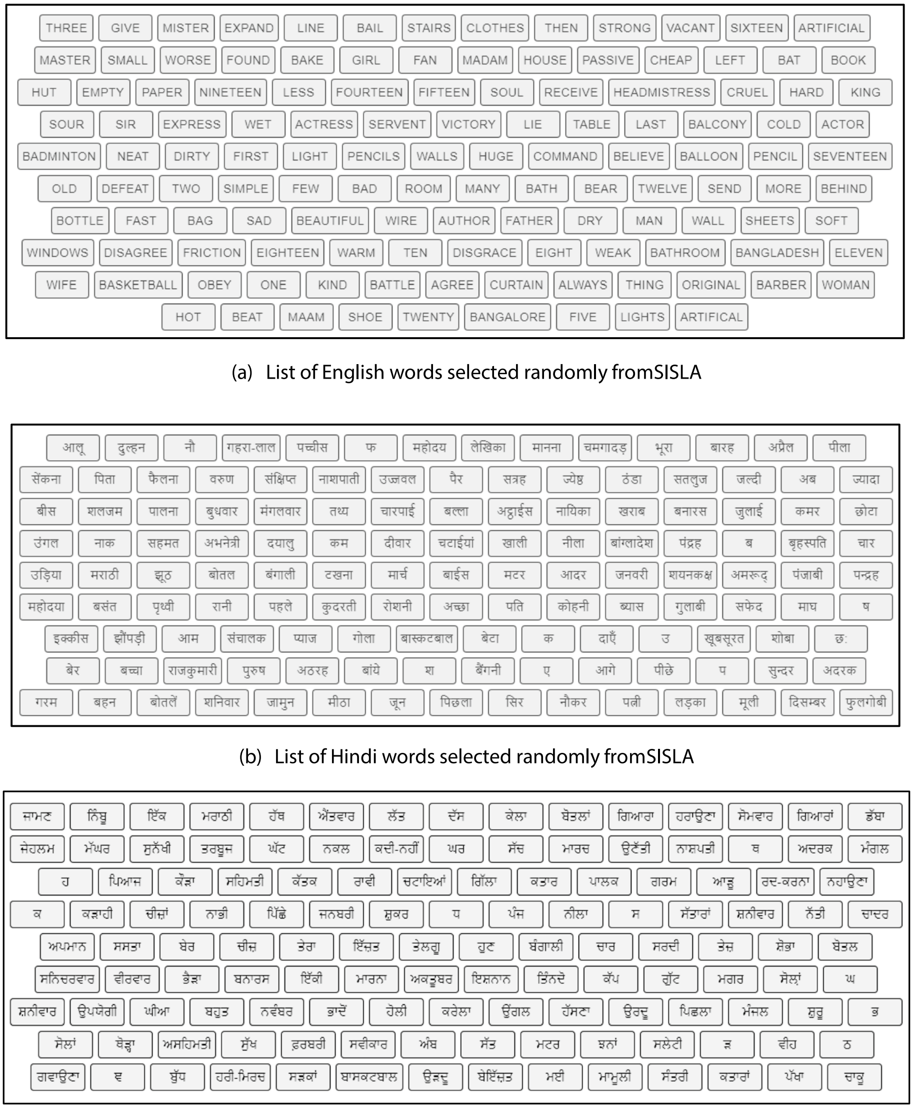
In the second part of this section, we have presented the sample(s) taken from 425 unique words selected from the following categories: Calendar, Family and Relation, Food, Countable, Animals, Health, Colors, Religion, Education, Emotions and Feelings, Sports, Interrogation, Popular Items, Other Words. The figures (Fig. 7a–c) displays the random words selected for the implementation of proposed system and their corresponding ISL synthetic corpus is prepared using HamNoSys.

6.3 Technological development

A multi-lingual ASR system, language translator and synthetic animation scheme have been integrated into a web-based application. SISLA has the feature to accept input in two types: (i) multi-lingual speech and (ii) multi-lingual text. The system has been developed

Table 2 Description of audio data prepared in this study

Parameters	Values
Audio file format	Extension: .wav extension Bitrate: 16 bits Sample rate: 16 kHz Channel: Single (Mono)
Number of Speakers	100 (50 Male, 50 Female)
Age Group	6 – 55 years
Number of Total Samples	16,878 English 16,821 Hindi 16,793 Punjabi
Average file size	200 KB
Average file length	6 Seconds
Corpus Size in Memory	~ 3 GB
Corpus Size in Time	~ 28 Hours



THREE GIVE MISTER EXPAND LINE BAIL STAIRS CLOTHES THEN STRONG VACANT SIXTEEN ARTIFICIAL

MASTER SMALL WORSE FOUND BAKE GIRL FAN MADAM HOUSE PASSIVE CHEAP LEFT BAT BOOK

HUT EMPTY PAPER NINETEEN LESS FOURTEEN FIFTEEN SOUL RECEIVE HEADMISTRESS CRUEL HARD KING

SOUR SIR EXPRESS WET ACTRESS SERVENT VICTORY LIE TABLE LAST BALCONY COLD ACTOR

BADMINTON NEAT DIRTY FIRST LIGHT PENCILS WALLS HUGE COMMAND BELIEVE BALLOON PENCIL SEVENTEEN

OLD DEFEAT TWO SIMPLE FEW BAD ROOM MANY BATH BEAR TWELVE SEND MORE BEHIND

BOTTLE FAST BAG SAD BEAUTIFUL WIRE AUTHOR FATHER DRY MAN WALL SHEETS SOFT

WINDOWS DISAGREE FRICTION EIGHTEEN WARM TEN DISGRACE EIGHT WEAK BATHROOM BANGLADESH ELEVEN

WIFE BASKETBALL OBEY ONE KIND BATTLE AGREE CURTAIN ALWAYS THING ORIGINAL BARBER WOMAN

HOT BEAT MAAM SHOE TWENTY BANGALORE FIVE LIGHTS ARTIFICAL

(a) List of English words selected randomly fromSISLA

आलू दुल्हन नौ गहरा-लाल पच्चीस फ महोदय लेखिका मानना चमगादड़ भूरा बारह अप्रैल पीला

सेंकना पिता फैलना वरुण संक्षिप्त नाशपाती उज्जवल पैर सत्रह ज्येष्ठ ठंडा सतलुज जल्दी अब ज्यादा

बीस शलजम पालना बुधवार मंगलवार तथ्य चारपाई बल्ला अट्ठाईस नायिका खराब बनारस जुलाई कमर छोटा

उंगल नाक सहमत अभनेत्री दयालु कम दीवार चटाईयां खाली नीला बांग्लादेश पंद्रह ब बृहस्पति चार

उड़िया मराठी झूठ बोतल बंगाली टखना मार्च बाईस मटर आदर जनवरी शयनकक्ष अमरूद् पंजाबी पन्द्रह

महोदया बसंत पृथ्वी रानी पहले कुदरती रोशनी अच्छा पति कोहनी ब्यास गुलाबी सफेद माघ ष

इक्कीस झौंपड़ी आम संचालक प्याज गोला बास्कटबाल बेटा क दाएँ उ खूबसूरत शोबा छः

बेर बच्चा राजकुमारी पुरुष अठरह बांये श बैंगनी ए आगे पीछे प सुन्दर अदरक

गरम बहन बोतलें शनिवार जामुन मीठा जून पिछला सिर नौकर पत्नी लड़का मूली दिसम्बर फुलगोबी

(b) List of Hindi words selected randomly fromSISLA

ਜਾਮਣ ਨਿੰਬੂ ਇੱਕ ਮਰਾਠੀ ਹੱਥ ਐਂਤਵਾਰ ਲੱਤ ਦੱਸ ਕੇਲਾ ਬੋਤਲਾਂ ਗਿਆਰਾ ਹਰਾਉਣਾ ਸੋਮਵਾਰ ਗਿਆਰਾਂ ਡੱਬਾ

ਜੇਹਲਮ ਮੱਘਰ ਸੁਨੱਖੀ ਤਰਬੂਜ ਘੱਟ ਨਕਲ ਕਦੀ-ਨਹੀਂ ਘਰ ਸੱਚ ਮਾਰਚ ਉਣੱਤੀ ਨਾਸ਼ਪਤੀ ਥ ਅਦਰਕ ਮੰਗਲ

ਹ ਪਿਆਜ ਕੌੜਾ ਸਹਿਮਤੀ ਕੱਤਕ ਰਾਵੀ ਚਟਾਇਆਂ ਗਿੱਲਾ ਕਤਾਰ ਪਾਲਕ ਗਰਮ ਆੜੂ ਰਦ-ਕਰਨਾ ਨਹਾਉਣਾ

ਕ ਕੜਾਹੀ ਚੀਜ਼ਾਂ ਨਾਭੀ ਪਿੱਛੇ ਜਨਬਰੀ ਸ਼ੁਕਰ ਧ ਪੰਜ ਨੀਲਾ ਸ ਸੱਤਾਰਾਂ ਸ਼ਨੀਵਾਰ ਨੱਤੀ ਚਾਦਰ

ਅਪਮਾਨ ਸਸਤਾ ਬੇਰ ਚੀਜ਼ ਤੇਰਾ ਇੱਜ਼ਤ ਤੇਲਗੂ ਹੁਣ ਬੰਗਾਲੀ ਚਾਰ ਸਰਦੀ ਤੇਜ਼ ਸ਼ੋਭਾ ਬੋਤਲ

ਸਨਿਚਰਵਾਰ ਵੀਰਵਾਰ ਭੈੜਾ ਬਨਾਰਸ ਇੱਕੀ ਮਾਰਨਾ ਅਕਤੂਬਰ ਇਸ਼ਨਾਨ ਤਿੰਨਦੋ ਕੱਪ ਗੁੱਟ ਮਗਰ ਸੋਲ਼ਾਂ ਘ

ਸ਼ਨੀਵਾਰ ਉਪਯੋਗੀ ਘੀਆ ਬਹੁਤ ਨਵੰਬਰ ਭਾਦੋਂ ਹੋਲੀ ਕਰੇਲਾ ਉਂਗਲ ਹੱਸਣਾ ਉਰਦੂ ਪਿਛਲਾ ਮੰਜਲ ਸ਼ੁਰੂ ਭ

ਸੋਲਾਂ ਥੋੜ੍ਹਾ ਅਸਹਿਮਤੀ ਸੁੱਖ ਫ਼ਰਬਰੀ ਸਵੀਕਾਰ ਅੰਬ ਸੱਤ ਮਟਰ ਝਨਾਂ ਸਲੇਟੀ ੜ ਵੀਹ ਠ

ਗਵਾਉਣਾ ਦ ਬੁੱਧ ਹਰੀ-ਮਿਰਚ ਸੜਕਾਂ ਬਾਸਕਟਬਾਲ ਉਜ਼ਦੂ ਬੇਇੱਜ਼ਤ ਮਈ ਮਾਮੂਲੀ ਸੰਤਰੀ ਕਤਾਰਾਂ ਪੱਖਾ ਚਾਕੂ

(c) List of Punjabi words selected randomly fromSISLA

Fig. 7 **a** List of English words selected randomly from SISLA. **b** List of Hindi words selected randomly from SISLA. **c** List of Punjabi words selected randomly from SISLA

using ASP.net along with the C# programming environment. In the proposed algorithm 1, an audio file has been used as the input data and then initialize stream speech recognizer to get result in an array. The stream speech recognizer translates the audio data to bit-stream and then decodes the audio data to textual results. A recognizer array stores the results provided by the speech recognizer. In the end of algorithm, plain text results have been stored in the resultant array. The proposed algorithm for speech to text conversion is listed below:

Algorithm 1 Speech to Text Conversion

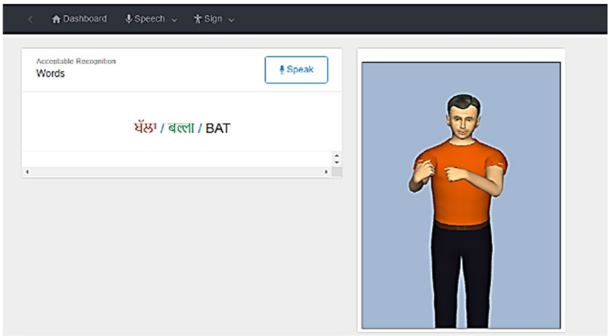
```
Input: AudioFile
Output: ResultArray
READ (AudioFile)
INITIALIZE (StreamSpeechReconizer)
INITIALIZE (RecognizedArray)
START (StreamSpeechReconizer)
GET RecognizedArray
If RecognizedArray NOT NULL
    for i = 0 to RecognizedArray.length
        PUSH into ResultArray
for i=0 to ResultArray.length
    display ResultArray[i]
```

6.3.1 User Interface

SISLA is a multi-lingual ASR where translation into an ISL system has been developed for performing the input and output operations in a user-friendly manner. Figure 8 displays the implemented SISLA system supporting English, Punjabi and Hindi language as input and generates 3D animation of ISL using HamNoSys script.

The developed system allows the user to perform the following operations:

- *Choice of input language:* English or Punjabi or Hindi.
- Speech Recognition: Start recognition, automatic stop recognition after 2 Seconds of time span.



Acceptable Recognition
Words

ਬੱਲਾ / बल्ला / BAT

Fig. 8 Implementation view of SISLA

- *Text Input:* English, Punjabi, Hindi Text Typing, filtration of the corpus, conversion into ISL.
- *Word selection*: Any isolated word at run-time selected and its corresponding gesture is displayed.
- *Multi-Lingual output:* Input language translates into English, Punjab, Hindi as well as in ISL.
- *Replay option* for ISL synthetic gesture.
- *Optional content:* Additional information of ISL video and image gesture is available.

6.3.2 Multi-lingual automatic speech recognition

In this paper, context-dependent speech recognition models have been trained using the CMU Sphinx toolkit. With the help of the CMU Sphinx toolkit, MFCC features have been extracted from speech data and the acoustic models have been generated based on extracted features. Figure 9 displays the process of speech recognition system using CMU Sphinx toolkit. Important files required for speech recognition model are:

a. ***Transcription files:*** Transcription file contains the textual content information about each audio sample file. Two separate transcription files for training and testing samples have been created with ". transcription" extension. Every spoken word in audio has been written in plain text along with a unique file id. Single space is required to separate each word and all the words have been enclosed in "<s>" and "</s>" tag as displayed in Table 3.

b) ***Audio File IDs List:*** Two separate files are required with ". fileids" extension, one for the training data set and second for the testing data set. Field-IDs must match with the

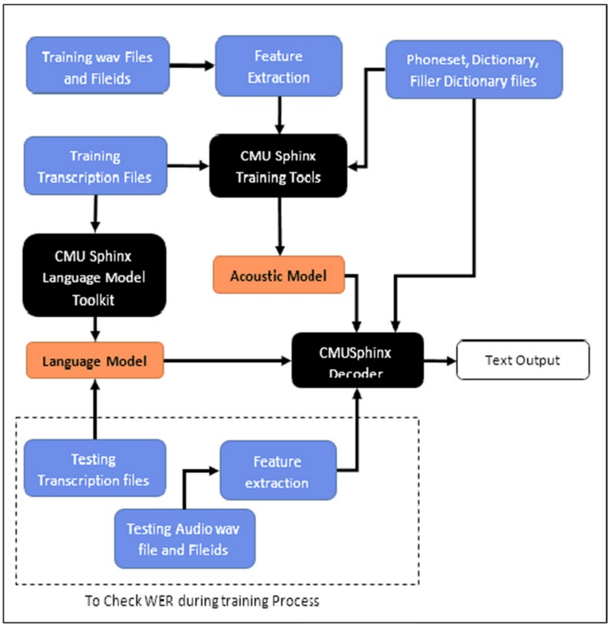

Training wav Files and Fileids
Feature Extraction
Phoneset, Dictionary, Filler Dictionary files
Training Transcription Files
CMU Sphinx Training Tools
CMU Sphinx Language Model Toolkit
Acoustic Model
Language Model
CMUSphinx Decoder
Text Output
Testing Transcription files
Feature extraction
Testing Audio wav file and Fileids
To Check WER during training Process

Fig. 9 Speech recognition process using CMU sphinx

Table 3 Samples of transcription files used during speech recognition training

Transcription file examples	Language
<s> ONE TWO THREE FIVE </s> (DPL-ID1-L1-S97) <s> ONE TWO THREE FIVE </s> (DPL-ID1-L1-S98) <s> ONE TWO THREE FOUR FIVE </s> (DPL-ID1-L1-S99) <s> ONE TWO THREE FOUR FIVE </s> (DPL-ID1-L1-S100)	English
<s> ਭਰਿਆ ਖਾਲੀ ਔਖਾ ਸੌਖਾ ਮਾੜਾ </s> (AD-ID2-L2-S47) <s> ਭਰਿਆ ਖਾਲੀ ਔਖਾ ਸੌਖਾ ਮਾੜਾ </s> (AD-ID2-L2-S48) <s> ਭਰਿਆ ਖਾਲੀ ਔਖਾ ਸੌਖਾ ਮਾੜਾ </s> (AD-ID2-L2-S49) <s> ਭਰਿਆ ਖਾਲੀ ਔਖਾ ਸੌਖਾ ਮਾੜਾ </s> (AD-ID2-L2-S50)	Punjabi
<s> केला अंगूर आम संतरा पपीता </s> (RJNT-ID15-L1-S47) <s> केला अंगूर आम संतरा पपीता </s> (RJNT-ID15-L1-S48) <s> केला अंगूर आम संतरा पपीता </s> (RJNT-ID15-L1-S49) <s> केला अंगूर आम संतरा पपीता </s> (RJNT-ID15-L1-S50)	Hindi

Table 4 File and directory structure for training process

Entries stored in IDs file as:
Example: • DPL/ID1-L1-S/DPL-ID1-L1-S98
Explanation: • “DPL” is a directory which is inside “wav” folder • “ID1-L1-S” is sub-directory of “DPL” directory. “ID1” contains the words that belong to the same category. “L1” means line ordering of the same group of words. “S” stands for speakers • DPL-ID1-L1-S98 is the recording file of 98th speaker.

sequence order in the transcription file as well as with the audio file, named along with the folder path as described in Table 4 inside the wav folder of CMU Sphinx toolkit.

Important points need consideration during file naming:

- The file name is case sensitive, so Python generates errors if file IDs are not matched properly
- No white space and/or special characters in the filename
- File names should be unique and meaningful

iii) **Phonetic Dictionary**: Phonetic dictionary holds every unique word which is in the transcription files. Every word is broken into multiple phonetic symbols based upon how a word is pronounced. Speech is segmented into several tones that are captured in a digital format.

There are various phoneset such as International Phonetic Alphabet (IPA) or Speech Assessment Methods Phonetic Alphabet (SAMPA) which are used to represent phonetic information about spoken word. CMU Sphinx is not limited only to the above mentioned phoneset, rather it also supports Unicode character sets. As an illustration, Table 5,

Table 5 Phonetic description of words in English, Punjabi and Hindi Language

Language	*Word*	*Phonetic description*
English	FOUR	F AO R
	FOURTEEN	F AO R T IY N
	NINE	N AY N
	NINETEEN	N AY N T IY N
Punjabi	ਅਠਾਰਾਂ	ਅ ਠਾ ਰਾਂ
	ਅਠਾਹਰਾ	ਅ ਠਾ ਹ ਰਾ
	ਅਦਰਕ	ਅ ਦ ਰ ਕ
	ਅਪਮਾਨ	ਅ ਪ ਮਾ ਨ
Hindi	अंतिम	अं ति म
	अकतूबर	अ क तू ब र
	अगस्‌त	अ ग स त
	अच्‌छा	अ च छा

Table 6 Sample of phoneset required for ASR

English	Punjabi	Hindi
AH	ੳ	आ
AO	ਓ	ई
AY	ਅ	उ
EH	ੲ	ए
ER	ਸ	क
EY	ਸ਼	ख
F	ਹ	ग
IH	ਕ	घ

describes the “FOUR” speech sound is broken into multiple speeches sounds as “F”, “AO”, “R”. The same approach is followed for Hindi and Punjabi language.

iv) **Phoneset file:** All the phonetic units are stored without any repetition in phoneset file along with “SIL” as silence word as shown in Table 6.

e) **Acoustic Model:** A special toolkit is used to construct an acoustic model as statistical representation of each word in audio recordings and their text transcriptions files. There are three types of acoustic model which CMU Sphinx support: continuous, semi-continuous and phonetically tied [32, 76, 86, 100]. For a larger dataset, the continuous acoustic model is used but these models require high processing power to decode speech signals. SISLA is configured and developed with a continuous acoustic model which is flexible to scale up the corpus.

f) **Language Model:** One more key component of ASR is the Language model. It provides assistance to the decoder to identify a sequence of words that are possible to recognize. The language model is created using *lmtool* from the online source provided by CMU Sphinx. It requires only plain text file which has all the sentences that the decoder has to recognize.

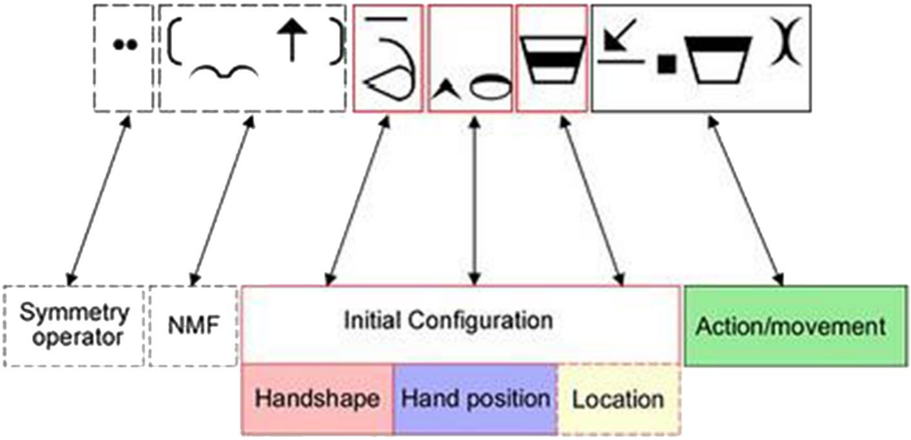


Fig. 10 Structure of HamNoSys Notation system

6.3.3 Sign language script conversion

HamNoSys is the most commonly used notation system for the research work of the text or speech to sign language machine translation. It supports the written notation system for sign language with the help of e-sign editor tool and the 3D virtual human avatar in JAService player java application. Generation of HamNoSys script for sign gestures require full concentration because HamNoSys also has its own structure that is displayed in Fig. 10. HamNoSys notation structure for single hand gesture consists of non-manual features, the shape of the hand, the orientation as well as location of the hand and optional features for the hand movement.

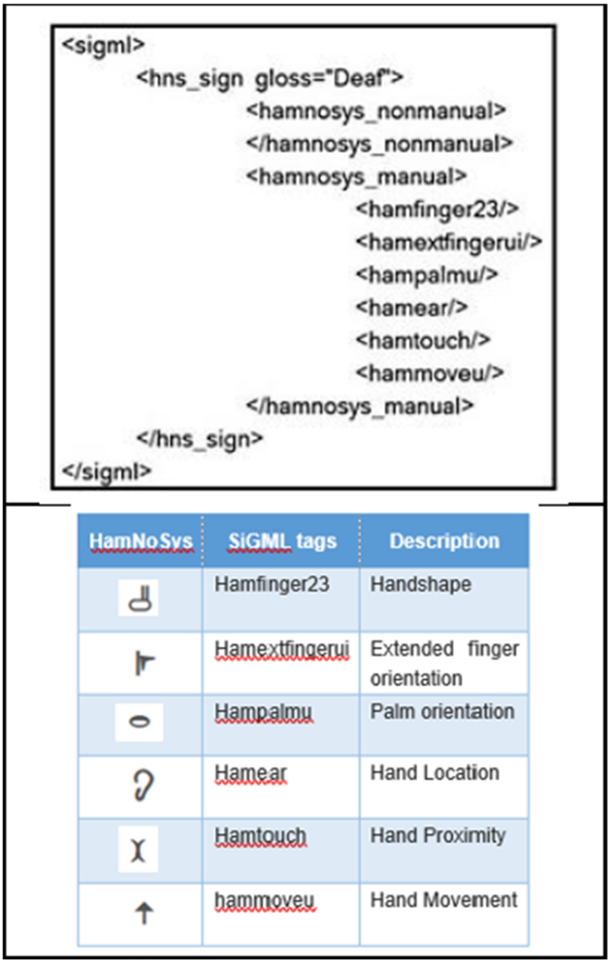
For both hands, an extra parameter such as symmetric operator is included in the beginning of HamNoSys notation. The symmetric operator defines whether the dominant hand copies the description to the non-dominant hand or defines differently [33, 46].

6.3.4 Markup document generation

HamNoSys notation is further translated into XML format which is known as SiGML notation [46, 47, 96, 113]. JASigning application takes input as SiGML scripted file and generates a sequence of synthetic signs in the form of a 3D avatar to the given word. Both manual and non-manual features are written using a sequence of corresponding tags in the SiGML script file. Figure 11 describes the SiGML file for the word “Deaf” along with the mapping of HamNoSys notation to SiGML tag with their description.

6.3.5 Synthetic animation

Generating signs using video approach or static animation is expensive and sometimes essential information of sign gesture is lost due to the 2D scene view. For example, in the gesture of {I, मै, मैं} signer’s index finger of the dominant hand point to oneself and rest of the fingers with the thumb are closed. So, it is difficult to estimate the angle of the index finger because it is hidden behind the rest of the hand part. Thus, synthetic animation is the best solution [12, 47, 51] for such kind of problems.



```
<sigml>
        <hns_sign  gloss="Deaf">
                    <hamnosys_nonmanual>
                    </hamnosys_nonmanual>
                    <hamnosys_manual>
                                <hamfinger23/>
                                <hamextfingerui/>
                                <hampalmu/>
                                <hamear/>
                                <hamtouch/>
                                <hammoveu/>
                    </hamnosys_manual>
        </hns_sign>
</sigml>
```

HamNoSys	SiGML tags	Description
	Hamfinger23	Handshape
	Hamextfingerui	Extended finger orientation
	Hampalmu	Palm orientation
	Hamear	Hand Location
	Hamtouch	Hand Proximity
↑	hammoveu	Hand Movement

Fig. 11 SiGML file structure and SiGML tags corresponding to HamNoSys notation symbols

6.4 Evaluation

SISLA has been developed to complete the objectives of research work which is to perform speech recognition of English, Punjabi and Hindi languages with further translation into ISL. To evaluate the developed system, quantitative and qualitative experiments have been performed with choice of certain parameters and conditions described in the next section.

6.4.1 Performance evaluation

This section describes the performance evaluation methods that are used to measure the quality of proposed research work. To evaluate the performance, three experiments are performed for the evaluation of speech recognition based on two factors: speaking environment (E1), speaking style (E2). Word Error Rate (WER) is measured with the use of confusion matrix based on these factors. E1 and E2 make use of randomly selected 200 from each input language.

In the first experiment, performance evaluation is done to test the impact of environmental factor on speech recognition. Experiment E1A has been performed inside a noiseless environment (i.e., in a closed room) and experiment E1B has been performed at a public place (in an academic institute). This experiment validates the interference effect of noise on the accuracy to recognize the spoken words. Speaking style varies in people according to their geographical area around the world. Dialects and fluency of speaking parameters also affect the speech recognition process which is observed in experiment E2. To test the second module of developed system, two experiments are designed related to the understanding of sign language evaluation test: E3 and Sign Error Rate (SER): E4.

To verify the sign accuracy, observations from experiment E3 are taken where a translation on the set of sign words is performed automatically. After that the sign language expert(s) manually checks the accuracy of sign generation system. The ISL experts evaluate the SISLA and provide the results based on correct and incorrect sign. In the final experiment E4, the outcome is analyzed for SISLA. Hearing people are considered for experiment-1, whereas hearing impaired users are selected for experiment-2. Hearing user speaks isolated word using microphone and hearing impaired user must judge the spoken word. Further, the selected users from both the categories (hearing impaired user with no prior knowledge of ISL and hearing impaired having good prior knowledge of ISL) are incredibly good in lipreading. One of them is selected to judge the spoken word based on lipreading and second user has been selected to judge using SISLA's 3D avatar only.

6.4.2 Results

This section discusses the results for performance and usability evaluation of the experiments. SISLA has been developed as a speech recognition system and achieving a higher accuracy was a major challenge due to its speaker independent nature. In the training process of speech model, the translator provided the word error rate (WER) of 7.3% (English), 9.9% (Punjabi) and 7.7% (Hindi) after the successful training model using CMU Sphinx toolkit. Furthermore, the implementation of these models in SISLA system, experiments E1 and E2 are performed using confusion matrix and the corresponding results are

Table 7 Experiment E1 results for the test parameter: Environment Noise

	English		Punjabi		Hindi	
	E1A-E	E1B-E	E1A-P	E1B-P	E1A-H	E1B-H
TN	21	29	24	38	23	40
FP	7	8	8	6	7	9
FN	8	11	12	17	10	14
TP	164	152	156	139	160	137
Total Inputs	**200**	**200**	**200**	**200**	**200**	**200**
Error-rate	0.07	0.11	0.10	0.11	0.08	0.11
Accuracy	0.93	0.91	0.90	0.89	0.92	0.89
Recall	0.95	0.93	0.93	0.89	0.94	0.91
Specificity	0.75	0.78	0.75	0.86	0.77	0.82
Precision	0.96	0.95	0.95	0.96	0.96	0.94

Table 8 Experiment E2 results for the test parameter: Speaking Style

	English		Punjabi		Hindi	
	E2A-E	E2B-E	E2A-P	E2B-P	E2A-H	E2B-H
TN	14	26	21	29	21	27
FP	5	7	9	6	7	9
FN	8	12	10	16	8	11
TP	173	155	160	149	164	153
Total Inputs	**200**	**200**	**200**	**200**	**200**	**200**
Error-rate	0.06	0.09	0.09	0.11	0.07	0.10
Accuracy	0.94	0.91	0.91	0.89	0.93	0.90
Recall	0.96	0.93	0.94	0.90	0.95	0.93
Specificity	0.74	0.79	0.70	0.83	0.75	0.75
Precision	0.97	0.95	0.95	0.96	0.96	0.94

displayed in Tables 7 and 8 respectively. E1A describes the outcomes for the experiment E1 conducted in the noiseless room and E1B for the noisy environment. The outcomes of the experiment are as follows:

- TP stands for True Positive: Word spoken correctly and recognized correctly
- TN stands for True Negative: Word spoken correctly and recognized incorrectly
- FP stands for False Positive: Word pronounced incorrectly but recognized correctly
- FN stands for False Negative: Word pronounced incorrectly and recognized incorrectly

Terms used in Table 7 for experiment-1 (Environment Noise):

- **E1A**: Experiment-1 performed in Noiseless environment
- **E1B**: Experiment-1 performed in less noisy environment
- **E**: English, **H**: Hindi and **P**: Punjabi

Parameters such as error-rate, accuracy, recall, specificity and precision are evaluated to show the goodness of system's performance where 1.0 is the best and 0.0 is the worst performance. WER is the number of words that are incorrectly recognized which are spoken properly or not. Overall, the WER for speech recognition of the developed system is low (7% (English), 10% (Punjabi) and 8% (Hindi)) for all the three input languages. For the noisy environment, experiment E1 gives the results as displayed in Table 7.

SISLA is working fine with an achieved accuracy of 91% for English, 89% for Punjabi and 90% for Hindi as compared to the accuracy achieved in the noiseless environment (English: 93%, Punjabi: 90%, Hindi: 92%). In this experiment, noiseless environment has signal to noise ratio, SNR $\geq$ 30 db. In order to improve the accuracy, spectral subtraction method or least mean square filter is suggested remove background noise effects. Use of multiple microphones also helps to capture area of interest and filter the background noise. To test the impact of speaking style on speech recognition system, experiments E2A and E2B are performed as shown in Table 8.

Terms used in Table 8 for experiment-2 (Speaking Style):

Table 9 ISL corpus results for the Experiment (E3)

Description	Words (%)
Perfectly generated	77%
Emotions extremely required	16%
Minor ISL parametric updates required	7%

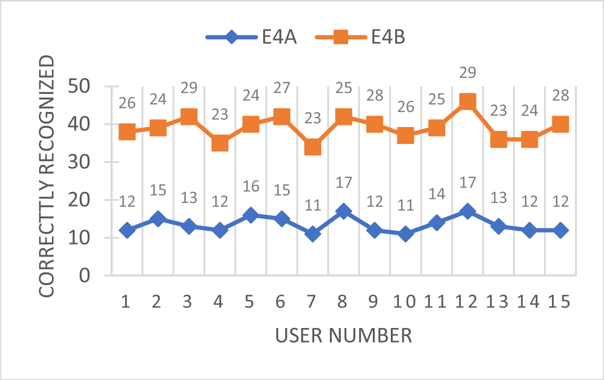


Fig. 12 Results for lipreading and sign gesture comparison

- **E2A**: Experiment-2 involved those speakers whose primary language is **E**: English, **H**: Hindi and **P**: Punjabi
- **E2B**: Experiment-2 involved those speakers which are able to speak E: English, H: Hindi and P: Punjabi but not fluently

E2A belongs to the test in which specific users have been selected with primary language as English, Hindi or Punjabi and are living in the Punjab region of India. In the other experiment E2B, the users have been selected that are able to speak in English, Hindi or Punjabi language but they do not use particular language as their primary language for communication. Developed system has shown impact on accuracy for those users which are not able to speak fluently in particular language as result displayed in Table 8. So, in SISLA machine translation system, users have choice to select their primary language from the available languages (English, Punjabi and Hindi).

The accuracy of ISL representation is observed in experiment (E3). Table 9 shows the result of developed systems that meets the phonetic feature of ISL gestures. ISL experts has noticed the issues of absence the non-manual features in developed synthetic sign gestures. Face expressions plays a vital role in sign language representation that include extra information about the feelings of signing person.

Currently developed system is not supporting the non-manual features but HamNoSys notation system supports the non-manual feature of sign language gesture so, in the future work it should be included for the improvement of SISLA performance. In the experiment (E4) both types of users have been selected to get the experimental results of the need of sign gestures as compared to lip reading. In the fig. E4A corresponds to the group of 15 users that guess the spoken words based only on lip reading. On the other hand, E4B, is the group of 15 different users that guess the spoken words based on ISL gestures. Figure 12 shows the experiment 4 results, E4A users judge less words correctly while E4B users judge higher words correctly. Experiment E4 results proven that sign language gestures

Table 10 Comparison of the proposed system with the existing systems

Ref.	Source Language	Target Language	Corpus	Speech	Speech Model	SL Notation	3D View Space	Output
[25]	Punjabi	ISL	100 words	✗	✗	✓	✓	Synthetic Animation
[73]	English	American Sign Language	–	✓	Existing API	✗	✗	Avatar Video
[41]	Kurdish	Kurdish Sign Language	560 words	✗	✗	✓	✓	Animation
[82]	Bangla	Bangla sign language	Bangla characters	✗	✗	✗	✗	Image
[105]	Arabic, English, German, French	African Sign Language	Not mentioned	✗	✗	✗	✗	Video
[91]	English	American Sign Language	600 words	✓	Model Build	✗	✗	–
[106]	English Hindi	Indian Sign Language	2000 words 3286 words	✗	✗	✓	✓	Animation
[65]	Arabic	Arabic Sign Language	3200 words	✗	✗	✗	✗	Video
[16]	Arabic	Arabic Sign Language	200 words	✗	✗	✓	✓	GIF Images
Our	**English** **Hindi** **Punjabi**	**Indian Sign Language**	**425 words** **413 words** **407 words**	✓ ✓ ✓	**Model Build**	✓	✓	**Animation**

contain more parameters to understand the spoken words and easy to remember as compared to lip reading. Terms used in Fig. 12 for the experiment-4 (E4) are listed below:

- **E4A**: A set of users for lip reading judgement
- **E4B**: A set of users for judgement of words based on sign gestures

Table 10 displays the comparison of proposed system in this study with the existing research work presented in past three year's studies. Most of the studies focused on the text input for the source language and few of them implemented the speech recognition systems. But in this study, the challenge to recognize the Punjabi and Hindi speech input has been taken into account because no public open license speech dataset is available. As already discussed in earlier sections of this study, synthetic sign language has many factors that provide better acceptance of gesture represent. But most of the authors have adopted video or static animation approach due to easiness and availability of sign language corpus. During this study it is observed that the neural network techniques are making impact to contribute in synthetic sign language but they are still at their initial stage of development. The only limitation of this system is that SISLA is developed to recognize isolated spoken words. ISL has its own grammar structure and rule to represent complete sentence. So, it requires more research work for the sentence level translation of spoken to sign language.

6.4.3 User acceptance evaluation

After use of SISLA, a user acceptance evaluation survey has been conducted (based on 5 points Likert scale) to provide the overall rate of satisfaction of the users. The presented proposed framework has been evaluated using 50 users: 25 experts are from the hearing impaired school and 25 are the family members of hearing impaired people. These users provide us the feedback about the automatic translation of speech to ISL. Each contributor

Table 11 User acceptance evaluation of the proposed framework for SISLA

The survey seeks to understand SISLA's performance as well as to make it more efficient. Please rate your experience and satisfaction level for SISLA.
Rate Score: (1 = Not Satisfactory, 2 = Somewhat Satisfactory, 3 = Neutral, 4 = Good, 5 = Excellent)

Sr. No.	Particulars	Options Score: 1 - 5
1.	SISLA is easy to use and learn a sign language	1 2 3 4 5
2.	SISLA fails to handle the request and crashes in between	1 2 3 4 5
3.	SISLA is an effective tool for learning and training in sign language	1 2 3 4 5
4.	I will recommend SISLA to my friends and family	1 2 3 4 5
5.	I am satisfied with the user interface of SISLA	1 2 3 4 5
6.	SISLA uses 3D characters to perform various gestures to make a better tool for understanding sign language	1 2 3 4 5
7.	Extra information provided with Signing avatar is beneficial	1 2 3 4 5
8.	Multi-lingual interaction system provides a better platform as compared to the use of single language	1 2 3 4 5
9.	Use of speech input is faster than typing text	1 2 3 4 5
10.	Overall SISLA is a beneficial tool as compared to other alternative tools.	1 2 3 4 5

Table 12 User acceptance average score results for SISLA

Sr. No.	Particulars	Average Score
1.	SISLA is easy to use and learn a sign language	3.9
2.	SISLA fails to handle the request and crashes in between	3.7
3.	SISLA is an effective tool for learning and training in sign language	4.4
4.	I will recommend SISLA to my friends and family	4.1
5.	I am satisfied with the user interface of SISLA	3.9
6.	SISLA uses 3D characters to perform various gestures to make a better tool for understanding sign language	4.5
7.	Extra information provided with Signing avatar is beneficial	4.1
8.	Multi-lingual interaction system provides a better platform as compared to the use of single language	4.8
9.	Use of speech input is faster than typing text	4.6
10.	Overall SISLA is a beneficial tool as compared to other alternative tools.	4.2

submits their response individually after the use of SISLA for at least 30 min. Table 11 list the various question included in the survey wizard filled by the participant.

User acceptance results have been collected using 5 points Likert scale are illustrated in Table 12. The table displays the average score of 50 users for each question in the survey. Users provide the highest average scores for the item 8 and item 9. SISLA gets the lower score for item 2, item 1 and item 5 which is lower than average score 4. Overall average score for all the questions in the Likert chart is 4.22 that shows SISLA is best tool for both hearing impaired as well as hearing people.

The automatic Machine translation system needs to recognize and parse the source language, translate into intermediate language and generate the result in the target language by graphical signing avatar. SISLA has three components: (i) trained model for speech recognizer, (ii) intermediate sign language translation module and (iii) representation into ISL. Combining these three components an internet-based frontend is developed with user-friendly options.

7 Discussion

Accuracy of the reference text transcription files has become more crucial factor for the remarkable accuracy of ASR. While preparing text transcription files for training and testing data, it is highly needed to clearly check and remove all the symbols that are irrelevant to speech signals. Following points lend a hand to decrease the error rate: (i) number format should be in written in source language script (English) instead of Numerical format (e.g., Source 5 should be written as Five), (ii) avoid to use acronym format (e.g., ASR word is pronounced completely as Automatic Speech Recognition in audio samples, so ASR acronym should increase the error rate), (iii) do not include any punctuation symbols in transcription file. Factors that affect the accuracy of the isolated speech recognition system are listed below:

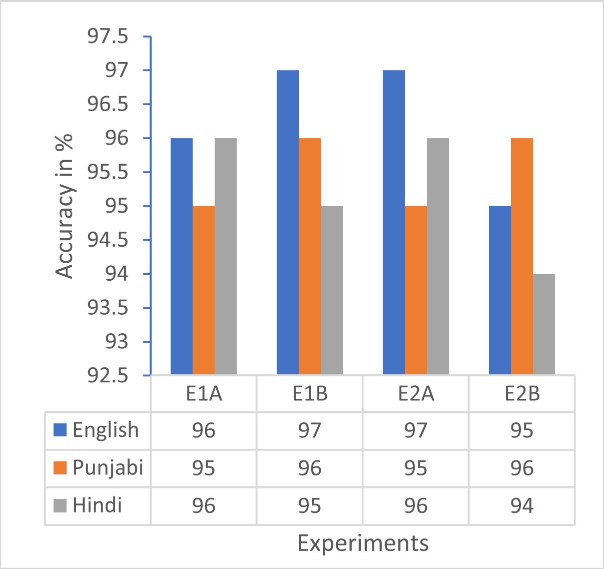
- Size of vocabulary (Number of words)
- Number of samples used to train the model

- Quality of data (speech as well as text transcription)
- Speaker variability (in terms of age and sex)
- Fluency of speaking depends upon the data used during training
- Background noise in speech samples
- Hardware computational power and microphone quality

Punjabi is a tonal language that means alteration in pitch results in a different word. It becomes difficult for the machine to recognize tonal language words. So, it is one of the reasons for lowest accuracy compared to Hindi and English. For example, ਕਰ (do) vs ਘਰ (house), ਕੜਾ (steel or iron bangle) vs ਘੜਾ (Earthen Pot), ਕਿਨਾ (how much) vs ਕਿਨ੍ਹਾਂ (who plural). Another reason for low accuracy of Hindi and Punjabi as compared to English is the number of phonemes used during training of ASR model for these languages. English model is trained with 425 vocabulary words that needed 43 phonemes for these words only. But training of ASR models for Hindi and Punjabi language requires a greater number of phonemes to extract and decode phonetic information. The vocabulary of 413 word for Hindi language requires 48 phonemes and 407 vocabulary word for Punjabi language require 54 phonemes. These are the technical parameters that are analyzed for the variation in accuracy of all the three speech recognition models. Overall accuracy of the presented system is illustrated in the Fig. 13 based on the experiments discussed in result section.

A common misconception about a sign language is that it is a universal language and it uses same gestures for particular word in all countries. In fact, like written or spoken languages, sign language also varies from country to country. But every sign language shares the same two main components: manual and non-manual features. There are 300 different sign languages around the globe [83]. Each sign language that the hearing impaired people use, belongs to particular social, cultural and/or religious groups of the country [47, 48, 70]. Reason of the diversity is that it develops naturally through the interaction of different groups of people. Due to variety of sign languages, it is difficult to develop a universal sign

Fig. 13 Proposed system's accuracy based on background noise and speaking style experiments



	E1A	E1B	E2A	E2B
English	96	97	97	95
Punjabi	95	96	95	96
Hindi	96	95	96	94

language. This research work focuses on the development of synthesis the ISL using notation system and 3D virtual avatar animation approach. Following key points show, how ISL is different from other sign languages as well as the dependency of spoken language.

a) Finger Spelling: When the sign for the particular word is unknown or word is a noun (such as name of place, people and item) or sign reader is unable to understand completely, then the finger spelling is used to perform sign language gestures. In finger spelling, the sequence of each alphabet exist in the sign word is performed on the basis of individual characters. Here, it is important to note that sign of the particular word and the finger spelling signs are always different.

Table 13, represents the difference of sign language gesture only for alphabets of spoken language in ISL. It is mandatory to create synthetic sign of every alphabet in all the input languages if system is designed to process input word along with finger spelling. In this study English, Hindi and Punjabi spoken languages are selected for the speech input. These languages have different number of letters for writing system: English 27 Alphabets, Hindi (33 alphabets and 11 Vowels) [101] and Punjabi (38 alphabets and 10 Vowels) [76]. The proposed system has included all the letters of these input languages and synthetic sign of each letter in ISL. This example shows the need of inclusion of each alphabetic representation. In the corpus, the word ("Loan", "ऋण", "ਕਰਜ਼ਾ") doesn't exist but the developed system is able to process the unknown word using finger spelling method for successful translation in ISL. If a word is considered as English input to the system, then system generate "L", "O", "A" and "N" gesture in ISL. If this word is considered as Hindi word, then "ऋ" and "ण" alphabets are used for finger spelling. Translation of Punjabi words is processed as the sequence of "ਕ", "ਰ", "ਜ਼" and "ਾ" using finger spelling method. But if the word is available in ISL corpus, then the complete word is mapped to the defined gesture rather than finger spell and the sign of that word should be same for all the three input languages (English, Hindi and Punjabi) in ISL.

b) Parametric difference with other sign languages: In ISL, single and both hands parameters are used to represent the sign gesture of alphabets. Whereas American Sign Language (ASL) use the single hand gestures and British Sign Language use both hands for all the English alphabets except for "C". Table 14 displays the ISL representation as compared to ASL and BSL for the English language alphabets (A, B, C, D and E). In this table input alphabets are same but they have different sign gestures in different sign languages.

Table 13 Difference of sign language gestures on basis of alphabets in English, Hindi and Punjabi language

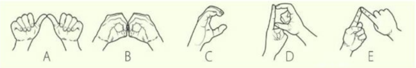
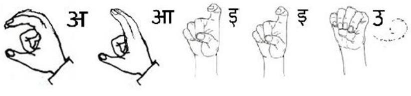
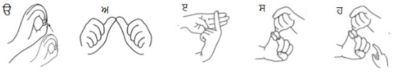
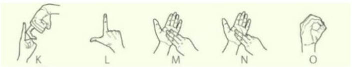
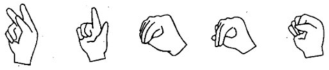
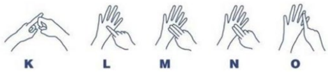
Spoken Language	Indian Sign language
English	
Hindi	
Punjabi	

Table 14 Comparison of sign languages for first five alphabets in English language

Sign Language Name	English Alphabet Representation
Indian Sign Language	
American Sign Language	
British Sign Language	

Because each country has its own sign language, a gesture has different meanings in different countries, as well as the same word is depicted using different gestures [24]. Several studies suggested that only the gesture recognition and movement of hands is not enough for correct translation in sign language [70].

ISL has different gestures for the same input word as compared to other sign languages such ASL, BSL and CSL. For example, sign gesture for the word “man” is performed as curling the moustache in ISL and then hand is bent, facing down, and moving above the right shoulder. But in British Sign Language, sign for same word is performed by setting up the shape of hand with thumb and fingers open around the chin and hand moves downward while thumb and fingers come together. BSL sign for man is closely related to the shape of beard. In ASL, open hand thump tip is contacted with forehead and moves downward to make a contact with chest in curvy shape. Same word “man” is differently performed in Chinese Sign Language with flat hand placing near to right side of head. So, it is clearly observed from the Fig. 14 that sign languages shares the common parameters to make a gesture but the gestures are not same for all the spoken languages.

iii) Sign language has also well-formed structure rules to make a meaningful sign gesture. For example, to point the person while showing who did what to whom, sign language uses a view space in front of the signer. However, verb sometimes refers to both the subject and the object, sometimes do not point at all and sometimes point only to object. Another rule of sign language is that well-formed question sentences must use eye

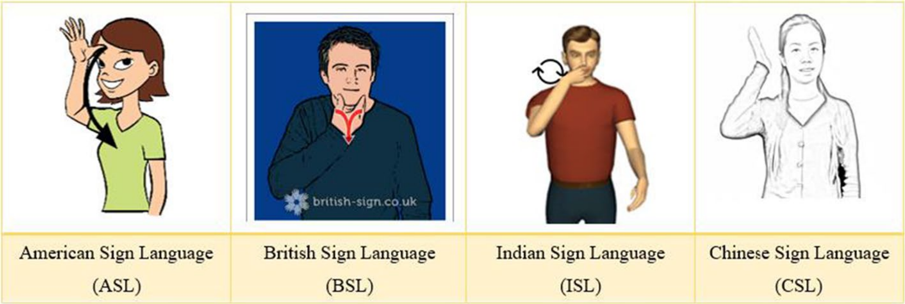

british-sign.co.uk
American Sign Language (ASL)
British Sign Language (BSL)
Indian Sign Language (ISL)
Chinese Sign Language (CSL)

Fig. 14 Difference of sign languages among them using example word “MAN” [15, 36, 55, 98]

expressions during sign gestures. Head movement is also important for better representation of yes/no sign gesture.

According to the studies and above-mentioned diversities in sign language, it is challenging task for the standardization of sign language.

8 Conclusion

Ultimately the aim of the research work is to support the first hypothesis: The development of SISLA as an automatic translator that translates multi-language speech into ISL using synthetic sign approach. The developed system supports spoken words in English. Punjabi and Hindi language as input and displays animated 3D avatar corresponding to the input word. During research work, various difficulties and challenges of the speech recognition system has been observed. Few of the challenges have been overcome with some precautions and with speech signal software applications. For example, during the collection of audio data for model training, pre-processing phase is applied through which the background noise is reduced with soundproof room for audio recording using better quality of the hardware (microphone). The silence between two words in the recording is carefully examined during recording for an isolated spoken word recognition system.

Its importance lies in the need as a tool for hearing impaired people of India that allows a fast and accurate translation system between speech and sign language. Speech recognition systems are used as human interaction input system without pressing keystrokes. These systems mainly work on the phonetic structure of sound. In speech recognition, audio data processed and analyzed based on feature vectors and the acoustic models helps to generate text corresponding to the data exists in feature vector. HamNoSys notation of the input word has been fetched from the ISL corpus which has been further translated into SiGML format. In the final stage, 3D avatar renders the gesture based on the SiGML file. Currently SISLA recognizes only isolated speech words for English, Hindi and Punjabi languages and translates them into ISL. But in future, further research work has to be done to recognize complete sentences and translate into ISL based on sign language grammar rules.

Acknowledgements This is supported by Department of Science and Technology, under the Scheme for Young Scientists & Technologists (SYST), by Government of India [SP/YO/382/2018(G)].

Funding This is supported by the Department of Science and Technology, under the Scheme for Young Scientists & Technologists (SYST), by the Government of India [SP/YO/382/2018(G)].

Declarations

Conflicts of interest/competing interests The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Abushariah MAM, Ainon RN, Zainuddin R, et al (2010) Natural speaker-independent Arabic speech recognition system based on hidden Markov models using Sphinx tools. In: international conference on computer and communication engineering (ICCCE'10). IEEE, pp 1–6
2. Ahmed M, Idrees M, Abideen Z ul, et al (2016) Deaf talk using 3D animated sign language: A sign language interpreter using Microsoft's kinect v2. In: 2016 SAI computing conference (SAI). IEEE, pp 330–335
3. Ahmed F, Bouillon P, Destefano C, et al (2017) Rapid construction of a web-enabled medical speech to sign language translator using recorded video. In: Quesada JF, Martín Mateos F-J, López Soto T (eds) Future and emerging trends in language technology. Machine learning and big data. FETLT 2016. Springer International Publishing, Cham, pp. 122–134
4. Alfi E, Atawy MS (2018) Intelligent Arabic sign language to Arabic text translation for easy deaf communication. International Journal of Computer Applications 180:19–26. https://doi.org/10.5120/ijca2018917081
5. Alkhalifa S, Al-Razgan M (2018) Enssat: wearable technology application for the deaf and hard of hearing. Multimed Tools Appl. Springer US, In, pp 1–25
6. Almasoud AM, Al-Khalifa HS (2012) SemSignWriting: a proposed semantic system for Arabic text-to-SignWriting translation. J Softw Eng Appl 05:604–612. https://doi.org/10.4236/jsea.2012.58069
7. Anuja K, Suryapriya S, Idicula SM (2009) Design and development of a frame based MT system for English-to-ISL. In: 2009 world congress on Nature & Biologically Inspired Computing (NaBIC). IEEE, pp 1382–1387
8. Arsan T, Ulgen O (2015) Sign language converter. International Journal of Computer Science & Engineering Survey 6:39–51. https://doi.org/10.5121/ijcses.2015.6403
9. Bernal Villamarin SC, Morales DAC, Reyes CAA, Sanchez CA (2016) Application design sign language colombian for mobile devices VLSCApp (voice Colombian sign language app) 1.0. In: 2016 technologies applied to electronics teaching (TAEE). IEEE, pp 1–5
10. Bhagwat SR, Bhavsar RP, Pawar B V. (2021) Translation from simple Marathi sentences to Indian sign language using phrase-based approach. In: 2021 international conference on emerging smart computing and informatics (ESCI). IEEE, pp 367–373
11. Bouzid Y, Jemni M (2013) An animated avatar to interpret signwriting transcription. International Conference on Electrical Engineering and Software Applications. IEEE, In, pp 1–5
12. Bouzid Y, Jemni M (2013) An avatar based approach for automatically interpreting a sign language notation. Proceedings - 2013 IEEE 13th international conference on advanced learning technologies, ICALT 2013 92–94. https://doi.org/10.1109/ICALT.2013.31
13. Bouzid Y, Jemni M (2014) TuniSigner: a virtual interpreter to learn sign writing. In: 14th international conference on advanced learning technologies. IEEE, pp 601–605
14. Bragg D, Caselli N, Gallagher JW et al (2021) ASL Sea Battle: Gamifying Sign Language Data Collection. In: Proceedings of the 2021 CHI conference on human factors in computing systems. ACM, New York, pp 1–13
15. British-sign.co.uk BSL-Sign (MAN). https://www.british-sign.co.uk/british-sign-language/how-to-sign/man/. Accessed 12 Aug 2021
16. Brour M, Benabbou A (2019) ATLASLang MTS 1: Arabic text language into Arabic sign language machine translation system. Procedia Computer Science 148:236–245. https://doi.org/10.1016/j.procs.2019.01.066
17. Bustamin A, Indrabayu, Areni IS, Mokobombang NN (2016) Speech to text for Indonesian homophone phrase with Mel Frequency Cepstral Coefficient. In: 2016 International Conference on Computational Intelligence and Cybernetics. IEEE, pp 29–31
18. Buttussi F, Chittaro L, Coppo M (2007) Using Web3D technologies for visualization and search of signs in an international sign language dictionary. In: proceedings of the twelfth international conference on 3D web technology - Web3D '07. ACM Press, New York, p 61
19. Caballero-Morales SO, Trujillo-Romero F (2013) 3D modeling of the Mexican sign language for a speech-to-sign language system. Computacion y Sistemas 17:593–608. https://doi.org/10.13053/CyS-17-4-2013-011
20. Chuan C, Guardino CA (2016) Designing SmartSignPlay: an interactive and intelligent American sign language app for children who are deaf or hard of hearing and their families. In: Companion publication of the 21st international conference on intelligent user interfaces - IUI '16 companion. ACM Press, New York, pp 45–48
21. Cox S, Lincoln M, Tryggvason J, et al (2002) TESSA, a system to aid communication with deaf people. In: the fifth international ACM conference on assistive technologies. ACM, pp 205–212

22. Das BP, Parekh R (2012) Recognition of Isolated Words using Features based on LPC , MFCC , ZCR and STE , with Neural Network Classifiers. International Journal of Modern Engineering Research (IJMER) 2:854–858
23. Debevc M, Kosec P, Holzinger A (2011) Improving multimodal web accessibility for deaf people: sign language interpreter module. Multimed Tools Appl 54:181–199. https://doi.org/10.1007/s11042-010-0529-8
24. Dewani A, Bhatti S, Memon MA et al (2018) Sign language e-learning system for hearing-impaired community of Pakistan. Int J Inf Technol 10:225–232. https://doi.org/10.1007/s41870-018-0105-4
25. Dhanjal A, Singh W (2020) An automatic conversion of Punjabi text to Indian sign language. ICST Transactions on Scalable Information Systems 7:1–10. https://doi.org/10.4108/eai.13-7-2018.165279
26. Elliott R, Glauert JRW, Kennaway JR et al (2008) Linguistic modelling and language-processing technologies for avatar-based sign language presentation. Univ Access Inf Soc 6:375–391. https://doi.org/10.1007/s10209-007-0102-z
27. Futane PR, Dharaskar R V. (2011) Hasta mudra an interpretation of Indian sign hand gestures. In: 2011 3rd international conference on electronics computer technology. IEEE, pp 377–380
28. Ghai W, Singh N (2013) Phone based acoustic modeling for automatic speech recognition for punjabi language. Journal of Speech Sciences 1:69–83. https://doi.org/10.1007/978-94-024-0846-1_100175
29. Glaser M, Tucker WD (2004) Telecommunications bridging between deaf and hearing users in South Africa. In: conference and workshop on assistive Technologies for Vision and Hearing Impairment (CVHI), pp 1–6
30. Gorman BM (2014) VisAural: a wearable sound-localisation device for people with impaired hearing. In: Proceedings of the 16th international ACM SIGACCESS conference on computers & accessibility. ACM, Rochester, pp 337–338
31. Grover Y, Aggarwal R, Sharma D, Gupta PK (2021) Sign language translation Systems for Hearing/speech impaired people: a review. In: 2021 international conference on innovative practices in technology and management (ICIPTM). IEEE, pp 10–14
32. Halawani SM, Daman D, Kari S, Ahmad AR (2013) An avatar based translation system from Arabic speech to Arabic sign language for deaf people. International Journal of Computer Science and Network Security 13:43–52
33. Hanke T (2004) HamNoSys – representing sign language data in language resources and language processing contexts. In: LREC. pp. 1–6
34. Hong R, Wang M, Xu M et al (2010) Dynamic captioning: video accessibility enhancement for hearing impairment. In: proceedings of the international conference on multimedia - MM '10. ACM Press, New York, p 421
35. Hong R, Wang M, Yuan X-T et al (2011) Video accessibility enhancement for hearing-impaired users. ACM Trans Multimed Comput Commun Appl 7S:1–19. https://doi.org/10.1145/2037676.2037681
36. IndianSignLanguage.org ISL-Sign Word (MAN). https://indiansignlanguage.org/man/. Accessed 12 Aug 2021
37. Jaballah K (2012) Accessible 3D signing avatars: the Tunisian experience. In: Proceedings of the international cross-disciplinary conference on web accessibility - W4A '12. ACM Press, New York, pp 1–2
38. Joy J, Balakrishnan K, Madhavankutty S (2018) Developing a bilingual mobile dictionary for Indian sign language and gathering users experience with SignDict. Assist Technol 32:153–160. https://doi.org/10.1080/10400435.2018.1508093
39. Kahlon NK, Singh W (2021) Machine translation from text to sign language: a systematic review. Univ Access Inf Soc:1–35. https://doi.org/10.1007/s10209-021-00823-1
40. Kallole NA, Prakash R (2018) Speech recognition system using open-source speech engine for Indian names. Lecture Notes in Electrical Engineering, In, pp 263–274
41. Kamal Z, Hassani H (2020) Towards Kurdish text to sign translation. In: Language Resources and Evaluation Conference. pp 117–122
42. Kaneko H, Hamaguchi N, Doke M, Inoue S (2010) Sign language animation using TVML. In: proceedings of the 9th ACM SIGGRAPH conference on virtual-reality continuum and its applications in industry - VRCAI '10. ACM Press, New York, p 289
43. Kar P, Reddy M, Mukerjee A, Raina AM (2007) INGIT : limited domain formulaic translation from Hindi strings to Indian sign language. In: International Conference on Natural Language Processing. pp. 1–10
44. Karpouzis K, Caridakis G, Fotinea SE, Efthimiou E (2007) Educational resources and implementation of a Greek sign language synthesis architecture. Computers and Education 49:54–74. https://doi.org/10.1016/j.compedu.2005.06.004

45. Karpov A, Kipyatkova I, Zelezny M (2016) Automatic Technologies for Processing Spoken Sign Languages. In: Procedia Computer Science. Elsevier, pp. 201–207
46. Kaur R, Kumar P (2014) HamNoSys generation system for sign language. In: International conference on advances in computing, Communications and Informatics. IEEE, pp. 2727–2734
47. Kaur K, Kumar P (2016) HamNoSys to SiGML conversion system for sign language automation. In: Procedia Computer Science. Elsevier, pp. 794–803
48. Kaur R, Kumar P (2017) Sign language based SMS generator for hearing impaired people. In: 2017 international conference on computational intelligence in data science(ICCIDS). IEEE, pp 1–5
49. Kaur S, Singh M (2015) Indian sign language animation generation system. In: 1st international conference on next generation computing technologies. IEEE, pp 909–914
50. Kaur R, Singh W (2018) Speech based retrieval system for Punjabi language. In: 2018 international conference on smart systems and inventive technology (ICSSIT). IEEE, pp 498–502
51. Kennaway JR, Glauert JRW, Zwitserlood I (2007) Providing signed content on the internet by synthesized animation. ACM Transactions on Computer-Human Interaction 14:1–29. https://doi.org/10.1145/1279700.1279705
52. Kheir R, Way T (2007) Inclusion of deaf students in computer science classes using real-time speech transcription. In: proceedings of the 12th annual SIGCSE conference on innovation and Technology in Computer Science Education. ACM, pp 261–265
53. Khilari P, P. B V. (2015) A review on speech to text conversion methods. Int J Adv Res Comput Eng Technol 4:3067–3072
54. Kipp M, Heloir A, Nguyen Q (2011) Sign language avatars: animation and comprehensibility. Intelligent Virtual Agents, In, pp 113–126
55. Kline A CSL-Sign Word (MAN). https://www.youtube.com/watch?v=cSxYv-3EMCc&t=1s. Accessed 12 Aug 2021
56. Kouremenos D, Fotinea SE, Efthimiou E, Ntalianis K (2010) A prototype Greek text to Greek sign language conversion system. Behaviour and Information Technology 29:467–481. https://doi.org/10.1080/01449290903420192
57. Kumar Y, Singh N (2017) An automatic speech recognition system for spontaneous Punjabi speech corpus. International Journal of Speech Technology 20:297–303. https://doi.org/10.1007/s10772-017-9408-2
58. Kumar K, Aggarwal RK, Jain A (2012) A Hindi speech recognition system for connected words using HTK. International Journal of Computational Systems Engineering 1:25–32. https://doi.org/10.1504/ijcsyse.2012.044740
59. Lee S, Kang Y, Lee Y (2016) LaneMate: Car sensing system for the deaf. In: Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems - CHI EA '16. ACM Press, New York, New York, USA, pp. 32–37
60. Lee S, Kang S, Han DK, Ko H (2016) Dialogue enabling speech-to-text user assistive agent system for hearing-impaired person. Medical & Biological Engineering & Computing 54:915–926. https://doi.org/10.1007/s11517-015-1447-8
61. López-Ludeña V, San-Segundo R, Lutfi S, et al (2011) Source language categorization for improving a speech into sign language translation system. In: Proceedings of the Second Workshop on Speech and Language Processing for Assistive Technologies. pp. 84–93
62. López-Zapata E, Campos-Trinidad M, Salazar-Arévalo R, Acostupa-Del Carpio J (2018) Design of a Robotic Speech-to-Sign Language Transliterating System. In: Carvalho JCM, Martins D, Simoni R, Simas H (eds) Mechanisms and machine science. Springer International Publishing, Cham, pp 274–282
63. Lu P, Huenerfauth M (2014) Collecting and evaluating the CUNY ASL corpus for research on American sign language animation. Comput Speech Lang 28:812–831. https://doi.org/10.1016/j.csl.2013.10.004
64. Luqman H, Mahmoud SA (2006) Automatic translation of Arabic text-to-Arabic sign language. ICGST Journal of Artificial Intelligence and Machine Learning 6:15–19
65. Luqman H, Mahmoud SA (2019) Automatic translation of Arabic text-to-Arabic sign language. Univ Access Inf Soc 18:939–951. https://doi.org/10.1007/s10209-018-0622-8
66. Mahmudul Hassan K, Hamid E, Molla KI (2017) A method for voiced/unvoiced classification of Noisy speech by analyzing time-domain features of spectrogram image. Science Journal of Circuits, Systems and Signal Processing 6, 11:–17. https://doi.org/10.11648/j.cssp.20170602.12
67. Malik M, Malik MK, Mehmood K, Makhdoom I (2021) Automatic speech recognition: a survey. Multimed Tools Appl 80:9411–9457. https://doi.org/10.1007/s11042-020-10073-7
68. Marschark M, Knoors H (2012) Educating deaf children: language, cognition, and learning. Deafness & Education International 14:136–160. https://doi.org/10.1179/1557069X12Y.0000000010

69. Martin PJM, Belhe S, Mudliar S, et al (2013) An Indian sign language (ISL) Corpus of the domain disaster message using avatar. In: proc. of the third international symposium in sign language translations and technology (SLTAT-2013), pp 1–4
70. Martins P, Rodrigues H, Rocha T, et al (2015) Accessible options for deaf people in e-learning platforms: technology solutions for sign language translation. In: 6th international conference on software development and Technologies for Enhancing Accessibility and Fighting Infoexclusion (DSAI 2015). Elsevier Masson SAS, pp 263–272
71. Matsumoto T, Kato M, Ikeda T (2009) JSPad - a sign language writing tool using SignWriting. In: proceedings of the 3rd international universal communication symposium, pp 363–367
72. Mean FO, Low TJ, La WW (2009) V2S: Voice to Sign Language Translation System for Malaysian Deaf People. In: International Visual Informatics Conference. IVIC 2009. In: Lecture notes in computer science. Springer, Berlin, Heidelberg, pp 868–876
73. Mehta N, Pai S, Singh S (2020) Automated 3D sign language caption generation for video. Univ Access Inf Soc 19:725–738. https://doi.org/10.1007/s10209-019-00668-9
74. Mirzaei MR, Ghorshi S, Mortazavi M (2012) Using augmented reality and automatic speech recognition techniques to help deaf and hard of hearing people. In: Proceedings of the 2012 virtual reality international conference on - VRIC ‘12. ACM Press, New York, pp 1–4
75. Mittal S, Kaur R (2016) Implementation of phonetic level speech recognition system for Punjabi language. In: 2016 1st India international conference on information processing (IICIP). IEEE, pp 1–6
76. Mittal P, Singh N (2019) Development and analysis of Punjabi ASR system for mobile phones under different acoustic models. International Journal of Speech Technology 22:219–230. https://doi.org/10.1007/s10772-019-09593-x
77. Mohandes M, Aliyu S, Deriche M (2015) Prototype Arabic sign language recognition using multisensor data fusion of two leap motion controllers. In: 2015 IEEE 12th international multi-conference on systems, Signals & Devices (SSD15). IEEE, pp 1–6
78. Mokhtar SA, Anuar SSS, Anuar SMS (2017) Web-based application for learning Malaysian sign language. In: Proceedings of the 11th international conference on ubiquitous information management and communication - IMCOM ‘17. ACM Press, New York, pp 1–6
79. Mon SM, Tun HM (2015) Speech-to-text conversion (STT) system using hidden Markov model (HMM). Int J Sci Technol Res 4:349–352
80. Morrissey S, Way A (2013) Manual labour: tackling machine translation for sign languages. Mach Transl 27:25–64. https://doi.org/10.1007/s10590-012-9133-1
81. Naert L, Larboulette C, Gibet S (2021) Motion synthesis and editing for the generation of new sign language content. Mach Transl. https://doi.org/10.1007/s10590-021-09268-y
82. Nawshin S, Saif N, Mohammad AS, Jameel M (2020) Protik: Bangla sign language teaching aid for children with impaired hearing. In: 2020 IEEE region 10 symposium (TENSYMP). IEEE, pp 440–443
83. Neves C, Coheur L, Nicolau H (2020) HamNoSyS2SiGML: translating HamNoSys into SiGML. In: LREC 2020 - 12th international conference on language resources and evaluation, conference proceedings, pp 6035–6039
84. Oliveira T, Escudeiro P, Escudeiro N, et al (2019) Automatic sign language translation to improve communication. In: 2019 IEEE global engineering education conference (EDUCON). IEEE, pp 937–942
85. Othman A, El Ghoul O, Jemni M (2010) SportSign: a service to make sports news accessible to deaf persons in sign languages. In: Computers helping people with special needs. Springer, Berlin, Heidelberg, pp 169–176
86. Otoom M, Alzubaidi MA (2018) Ambient intelligence framework for real-time speech-to-sign translation. Assist Technol 30:119–132. https://doi.org/10.1080/10400435.2016.1268218
87. Padmanabhan J, Johnson Premkumar MJ (2015) Machine learning in automatic speech recognition: a survey. IETE Tech Rev 32:240–251. https://doi.org/10.1080/02564602.2015.1010611
88. Parton BS (2006) Sign language recognition and translation: a multidisciplined approach from the field of artificial intelligence. J Deaf Stud Deaf Educ 11:94–101. https://doi.org/10.1093/deafed/enj003
89. Pražák A, Loose Z, Psutka JV et al (2020) Live TV subtitling through respeaking with remote cutting-edge technology. Multimed Tools Appl 79:1203–1220. https://doi.org/10.1007/s11042-019-08235-3
90. Priyanks BR, Mukund A (2019) Tamil speech to Indian sign language using CMUSphinx language models. International Research Journal of Engineering and Technology 6:1812–1814

91. Qaisar SM, Niyazi S, Subasi A (2019) Efficient isolated speech to sign conversion based on the adaptive rate processing. Procedia Computer Science 163:35–40. https://doi.org/10.1016/j.procs.2019.12.083
92. Reddy BR, Mahender E (2013) Speech to text conversion using android platform. International Journal of Engineering Research and Application 3:253–258
93. Sahu PK, Ganesh DS (2015) A study on automatic speech recognition toolkits. In: 2015 international conference on microwave, Optical and Communication Engineering (ICMOCE). IEEE, pp 365–368
94. Saifan RR, Dweik W, Abdel-Majeed M (2018) A machine learning based deaf assistance digital system. Comput Appl Eng Educ 26:1008–1019. https://doi.org/10.1002/cae.21952
95. Samčović A (2020) Accessibility of services in digital television for hearing impaired consumers. Assistive Technology:1–10. https://doi.org/10.1080/10400435.2020.1757786
96. San-Segundo R, Barra R, Córdoba R et al (2008) Speech to sign language translation system for Spanish. Speech Comm 50:1009–1020. https://doi.org/10.1016/j.specom.2008.02.001
97. Sarma H, Saharia N, Sharma U (2017) Development and analysis of speech recognition Systems for Assamese Language Using HTK. ACM Transactions on Asian and Low-Resource Language Information Processing 17:1–14. https://doi.org/10.1145/3137055
98. Savvy S ASL-Sign Word (Man). https://www.signingsavvy.com/sign/MAN. Accessed 12 Aug 2021
99. Sawant S, Deshpande M (2018) Isolated spoken Marathi words recognition using HMM. in: fourth international conference on computing communication control and automation (ICCUBEA). IEEE, pp 1–4
100. Shahriar R, Zaman AGM, Ahmed T, et al (2017) A communication platform between bangla and sign language. In: 5th IEEE region 10 humanitarian technology conference 2017, R10-HTC 2017, pp 1–4
101. Sharan S, Bansal S, Agrawal SS (2018) Speaker-independent recognition system for continuous Hindi speech using probabilistic model. In: Advances in Intelligent Systems and Computing. Springer, pp 91–97
102. Singh S, Deep Kaur P (2016) Subjective well-being prediction from social networks: a review. In: 2016 fourth international conference on parallel. Distributed and Grid Computing (PDGC), IEEE, pp 90–95
103. Singh P, Dutta K (2011) Formant analysis of Punjabi non-nasalized vowel phonemes. In: 2011 international conference on computational intelligence and communication networks. IEEE, pp 375–380
104. Sonawane P, Shah K, Patel P, et al (2021) Speech to Indian sign language (ISL) translation system. In: 2021 international conference on computing, communication, and intelligent systems (ICCCIS). IEEE, pp 92–96
105. Soudi A, Van Laerhoven K, Bou-Souf E (2019) AfricaSign -- a crowd-sourcing platform for the documentation of STEM vocabulary in African sign languages. In: The 21st international ACM SIGACCESS conference on computers and accessibility. ACM, New York, pp 658–660
106. Sugandhi KP, Kaur S (2018) Online multilingual dictionary using Hamburg notation for avatar-based Indian sign language generation system. International Journal of Cognitive and Language Sciences 12:1116–1122. https://doi.org/10.5281/zenodo.1474397
107. Syms S, Wang H (2021) Naive Bayes and Entropy based Analysis and Classification of Humans and Chat Bots. Journal of ISMAC 3:40–49. https://doi.org/10.36548/jismac.2021.1.004
108. Thimmaraja Yadava G, Jayanna HS (2020) Enhancements in automatic Kannada speech recognition system by background noise elimination and alternate acoustic modelling. International Journal of Speech Technology 23:149–167. https://doi.org/10.1007/s10772-020-09671-5
109. Tripathy S, Baranwal N, Nandi GC (2013) A MFCC based Hindi speech recognition technique using HTK toolkit. In: 2013 IEEE second international conference on image information processing (ICIIP-2013). IEEE, pp 539–544
110. Upadhyaya P, Mittal SK, Farooq O, et al (2019) Continuous Hindi speech recognition using Kaldi ASR based on Deep neural network. In: Advances in Intelligent Systems and Computing. Springer Singapore, pp 303–311
111. Valanarasu R (2021) Comparative Analysis for Personality Prediction by Digital Footprints in Social Media. Journal of Information Technology and Digital World 3:77–91. https://doi.org/10.36548/jitdw.2021.2.002
112. Van Zijl L, Combrink A (2006) The south African sign language machine translation project: issues on non-manual sign generation. In: Proceedings of the south African institute of computer scientists and information technologists on IT research in developing couuntries - SAICSIT ‘06. ACM Press, New York, pp 127–134
113. Varghese M, Nambiar SK (2018) English to SiGML conversion for sign language generation. In: 2018 international conference on circuits and Systems in Digital Enterprise Technology (ICCSDET). IEEE, pp 1–6

114. Verma A, Kaur S (2015) Indian sign language automation Genration system for Gurmukhi script. Int J Comput Sci Technol 6:117–121
115. Wray A, Cox S, Lincoln M, Tryggvason J (2004) A formulaic approach to translation at the post office: Reading the signs. Lang Commun 24:59–75. https://doi.org/10.1016/j.langcom.2003.08.001
116. Yao D, Qiu Y, Huang H (2009) Web-based Chinese sign language broadcasting system. Proceedings of the 2009 international cross-disciplinary conference on web accessibililty (W4A) 101–103. https://doi.org/10.1145/1535654.1535680
117. Zerari N, Yousfi B, Abdelhamid S (2016) Automatic speech recognition: a review. International Journal of Computer Applications 2:63–68. https://doi.org/10.5120/9722-4190

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Amandeep Singh Dhanjal is a research scholar of Ph.D. in Department of Computer Science in Punjabi University Patiala, Punjab, India, He has earned his Master of Computer Application in 2011 from Punjab Technical University Jalandhar, Punjab, India. He has already published some research papers in conferences. His areas of interest are Natural Language Processing, Speech Technology, Indian Sign Language, Software Development.

Williamjeet Singh is working as an Assistant Professor in Department of Computer Science and Engineering at Punjabi University, Patiala, Punjab, India. He achieved his BTech (CSE) degree from BBSBEC, Fatehgarh Sahib under Punjab Technical University in 2005. He completed his MTech (CSE) from Punjabi University, Patiala in the year 2007. He was awarded PhD degree in the year 2015 in the faculty of Engineering and Technology from Punjabi University. His areas of interest include Sign Language, Speech Recognition, Cellular Networks, Algorithms, Speech Technology, Data Mining and Sentiment Analysis. He has more than 11 years of teaching and research experience. He has published many research papers in conferences and journals of National as well as international repute.