

Article

# 3D Avatar Approach for Continuous Sign Movement Using Speech/Text

Debashis Das Chakladar <sup>1</sup>, Pradeep Kumar <sup>1</sup>, Shubham Mandal <sup>1</sup>, Partha Pratim Roy <sup>1</sup>, Masakazu Iwamura <sup>2</sup>  
and Byung-Gyu Kim <sup>3,\*</sup>

<sup>1</sup> Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, Uttarakhand 247667, India; dchakladar@cs.iitr.ac.in (D.D.C.); pradeep.iitr7@gmail.com (P.K.); shubhammandal96@gmail.com (S.M.); proy.fcs@iitr.ac.in (P.P.R.)

<sup>2</sup> Department of Computer Science and Intelligent Systems, Osaka Prefecture University, Osaka 599-8531, Japan; masa@cs.osakafu-u.ac.jp

<sup>3</sup> Department of IT Engineering, Sookmyung Women's University, Seoul 140-742, Korea

\* Correspondence: bg.kim@sookmyung.ac.kr

**Abstract:** Sign language is a visual language for communication used by hearing-impaired people with the help of hand and finger movements. Indian Sign Language (ISL) is a well-developed and standard way of communication for hearing-impaired people living in India. However, other people who use spoken language always face difficulty while communicating with a hearing-impaired person due to lack of sign language knowledge. In this study, we have developed a 3D avatar-based sign language learning system that converts the input speech/text into corresponding sign movements for ISL. The system consists of three modules. Initially, the input speech is converted into an English sentence. Then, that English sentence is converted into the corresponding ISL sentence using the Natural Language Processing (NLP) technique. Finally, the motion of the 3D avatar is defined based on the ISL sentence. The translation module achieves a 10.50 SER (Sign Error Rate) score.

**Keywords:** Indian Sign Language (ISL); natural language processing; avatar; sign movement; context-free grammar



**Citation:** Das Chakladar, D.; Kumar, P.; Mandal, S.; Roy, P.P.; Iwamura, M.; Kim, B.-G. 3D Avatar Approach for Continuous Sign Movement Using Speech/Text. *Appl. Sci.* **2021**, *11*, 3439. <https://doi.org/10.3390/app11083439>

Academic Editor: Andrea Prati

Received: 15 February 2021

Accepted: 2 April 2021

Published: 12 April 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

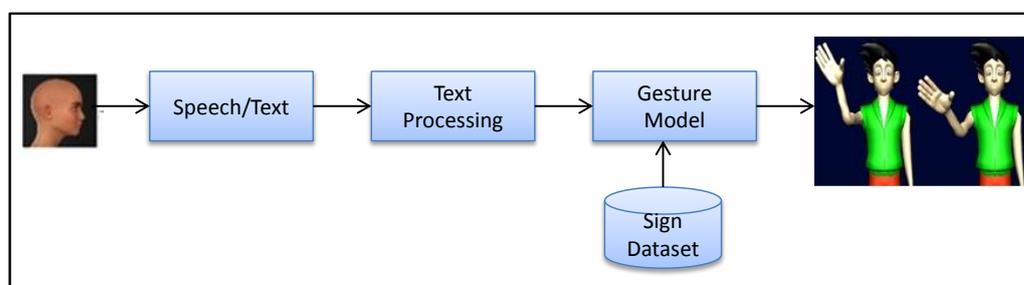
## 1. Introduction

A sign is a sequential or parallel construction of its manual and non-manual components. A manual component can be defined by hand shape, orientation, position, and movements, whereas non-manual components are defined by facial expressions, eye gaze, and head/body posture [1–5]. Hearing-impaired people use sign language for their communication. Every country has its sign language based on its vocabulary and syntax. Therefore, sign translation from speech/text is specific to the particular targeted country. Indian Sign Language (ISL) is one of the sign languages that can be efficiently translated from English. Moreover, ISL is recognized as a widely accepted natural language for its well-defined grammar, syntax, phonetics, and morphology structure over others [6]. ISL is a visual-spatial language that provides linguistic information using the hands, arms, face, and head/body postures. The ISL open lexicon can be categorized into three parts: (i) Signs whose place of articulation is fixed, (ii) signs whose place of articulation can change, and (iii) directional signs, where there is a movement between two points in space [7]. However, people who use English as a spoken language do not understand the ISL. Therefore, an English to ISL sign movement translation system is required for assistance and learning purposes.

In India, more than 1.5 million people are hearing-impaired who use ISL as their primary means of communication [8]. Some studies [6,8,9] implemented ISL videos for sign representation from English text. To generate a robust sign language learning system from

English to ISL, output sign representation should be efficient, such as being able to generate proper signs without delay for complete sentences. However, sign language translation from ISL video recordings requires notable processing time [6]. By contrast, the sign representations using a 3D avatar require minimum computational time, and the avatar can be easily reproduced as per the translation system [10]. Moreover, most of the existing studies [11,12] have not considered complete sentences for sign language conversion. To overcome these shortcomings, in this paper, we propose a 3D avatar-based ISL learning model that can perform sign movements not only for isolated words but also for complete sentences through input text or speech. The flow diagram of such an assisting system is depicted in Figure 1, where the input to the system is either English speech or text, which is then processed using a text processing technique to obtain ISL representation. Next, a gesture model is used to perform the sign movement corresponding to ISL with the help of an avatar. The main contributions of the work are defined as follows.

- Our first contribution is the development of a 3D avatar model for Indian Sign Language (ISL).
- The proposed 3D avatar model can generate sign movements from three different inputs, namely speech, text, and complete sentences. The complete sentence obtained is made up of continuous signs corresponding to a sentence of spoken language.



**Figure 1.** An assistive model for generating sign movements using the 3D avatar from English speech/text.

The rest of the paper is organized as follows. The related work of sign language translation is discussed in Section 2. In Section 3, we describe the proposed model of speech to sign movement for ISL sentences. In Section 4, we analyze each module of our proposed model. Finally, Section 5 presents the conclusion of this paper.

## 2. Related Work

This section consists of two subsections: sign language translation systems and performance analysis of the sign language translation systems. A detailed description of each module is given below.

### 2.1. Sign Language Translation Systems

Sign movement can be effectively generated from input speech. In [13], the authors have designed a speech–sign translation system for Spanish Sign Language (SSL) using a speech recognizer, a natural language translator, and a 3D avatar animation module. In [11,14], the authors have implemented the conversion of Arabic Sign Language (ArSL) from Arabic text using an Arabic text-to-sign translation system. The translation system uses the set of translation rules and linguistic language models for detecting different signs from the text. An American Sign Language (ASL)-based animation sequence has been proposed in [15]. The authors’ system converts all of the hand symbols and associated movements of the ASL sign box. A speech-to-sign movement translation based on Spanish Sign Language (SSL) has been proposed in [12]. The authors used two types of translation techniques (rule-based and statistical) of the Natural Language Processing (NLP) toolbox to generate SSL. A linguistic model-based 3D avatar (for British Sign Language) has been proposed for implementing the visual realization of sign language [16]. A web-based

interpreter from text to sign language was developed in [17]. The interpreter tool has been created from a large dictionary of ISL such that it can be shared among multilingual communities. An android app-based translation system has been designed to convert sign movements from hand gestures of ISL [18]. In [19], the authors designed a Malayalam text to ISL translation system using a synthetic animation approach. Their model has been used to promote sign language education among the common people of Kerala, India. A Hindi text to the ISL conversion system has been implemented in [20]. Their model used the dependency parser and Part-of-Speech (PoS) tagger, which correctly categorize the input words into their syntactic forms. An interactive 3D avatar-based math learning system of American Sign Language (ASL) has been proposed in [21]. The math-learner model can increase the effectiveness of parents of hearing-impaired children in teaching mathematics to their children. A brief description of existing sign language learning systems is presented in Table 1. It can be observed that some sign language translation models work on speech-to-sign conversion, whereas some models translate the text to signs and represent the signs using a 3D avatar. Our proposed model successfully converts the input speech to corresponding text and then renders the signs movements using the 3D avatar.

**Table 1.** Brief description of previous studies of sign language learning systems. Note: Arabic Sign Language (ArSL), Chinese Sign Language (CSL), Spanish Sign Language (SSL), American Sign Language (ASL), and Indian Sign Language (ISL). “Sentence-wise sign” represents the continuous signs corresponding to a sentence in its correspondent spoken language.

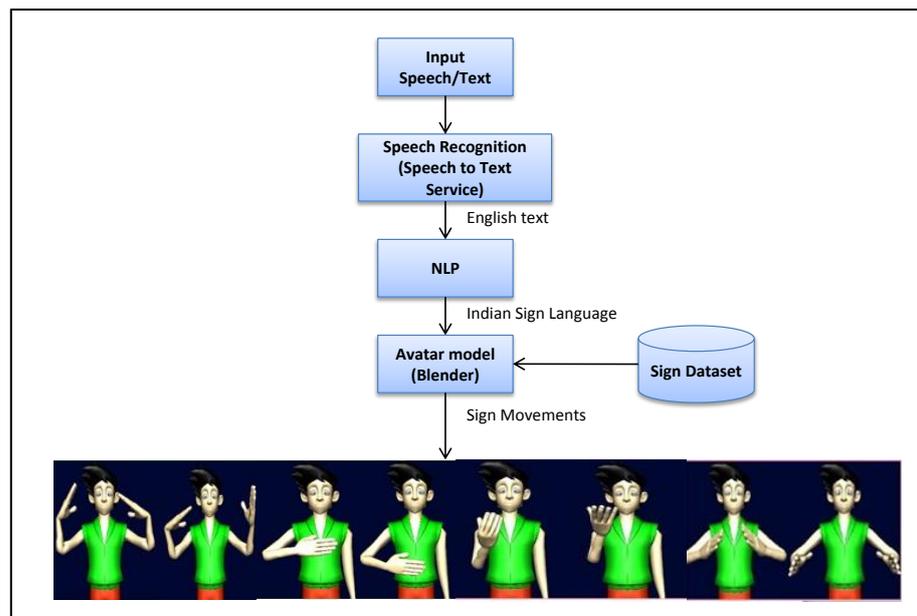
Study	Sign Language	Input: Speech	Input: Text	3D Avatar	Sentence-Wise Sign
Al-Barahamtoshy, O.H. et al. [11]	ArSL	✗	✓	✓	✗
Li et al. [22]	CSL	✓	✗	✗	✓
Halawani et al. [14]	ArSL	✓	✗	✗	✗
Lopez-Ludena et al. [23]	SSL	✗	✓	✓	✗
Bouzid, Y. et al. [24]	ASL	✗	✗	✓	✗
Dasgupta et al. [8]	ISL	✗	✓	✗	✗
Nair et al. [19]	ISL	✗	✓	✗	✓
Vij et al. [20]	ISL	✗	✓	✗	✗
Krishnaraj et al. [6]	ISL	✓	✓	✗	✗
Duarte et al. [25]	ASL	✓	✗	✗	✓
Patel et al. [26]	ISL	✓	✗	✓	✓
<b>Proposed</b>	ISL	✓	✓	✓	✓

## 2.2. Performance Analysis of the Sign Language Translation System

This section discussed the effectiveness of the different sign language translation systems based on different evaluation metrics such as Sign Error Rate (SER), Bilingual Evaluation Understudy (BLEU), and the National Institute of Standards and Technology (NIST). In [27], the authors have designed an avatar-based model that can generate sign movements from spoken language sentences. They achieved a 15.26 BLEU score with a Recurrent Neural Network (RNN)-based model. In [28,29], the authors have proposed a “HamNoSys” system that converts the input words to corresponding gestures of ISL. “HamNoSys” represents the syntactic representation of each sign using some symbols, which can be converted into the respective gestures (hand movement, palm orientation). Apart from “HamNoSys”, Signing Gesture Markup Language (SiGML) [30] also has been used for transforming sign visual representations into a symbolic design. In [31], the authors have used BLEU and the NIST score, which are relevant for performance analysis of language translation. Speech to SSL translation has been implemented with two types of natural language-based translations (rule-based and statistical) [12]. The authors have identified that rule-based translation outperforms statistical translation with a 31.6 SER score and a 0.5780 BLEU score.

### 3. Materials and Methods

This section illustrates the framework of the proposed sign language learning system for ISL. The proposed model is subdivided into three modules, as depicted in Figure 2. The first module corresponds to the conversion of speech to an English sentence, which is then processed using NLP to obtain the corresponding ISL sentence. Lastly, we feed the extracted ISL sentence to the avatar model to produce the respective sign language. We discuss the detailed description of each module in Sections 3.1–3.3.



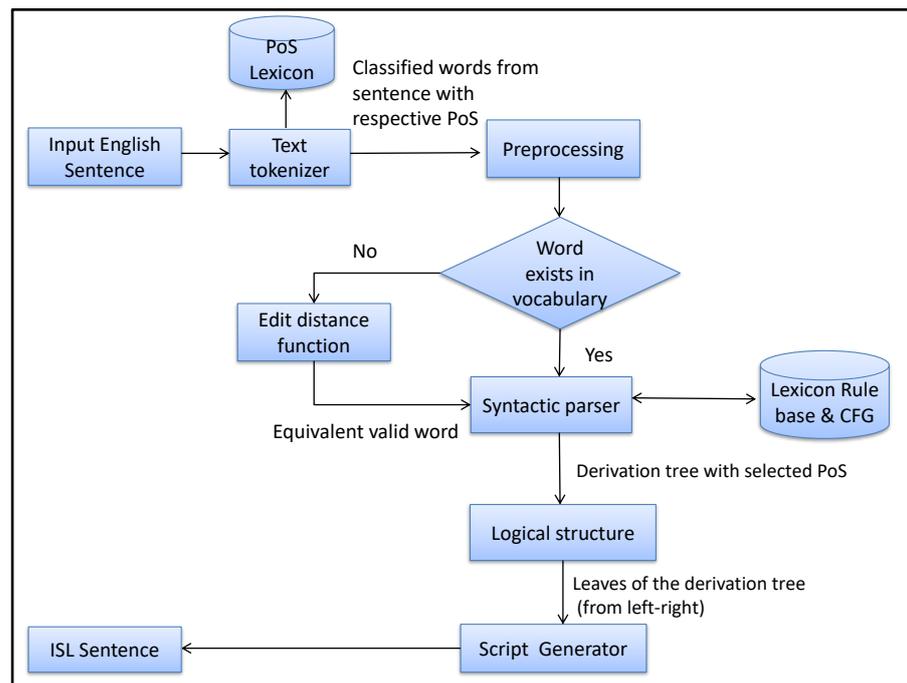
**Figure 2.** Framework of the proposed sign language learning system using text/speech. NLP: Natural Language Processing.

#### 3.1. Speech to English Sentence Conversion

We used the IBM-Watson service (available online: <https://cloud.ibm.com/apidocs/speech-to-text> (accessed on 20 December 2020)). to convert the input speech into text. The service is classified into three phases, i.e., input features, interfaces, and output features. The first phase illustrates the input audio format (.wav, .flac, .mp3, etc.) and settings (sampling rate, number of communication channels) of the speech signal. Next, an HTTP request is generated for each speech signal. The input speech signal interacts with the speech-to-text service using various interfaces (web socket interface, HTTP interface, and asynchronous HTTP interface) using the communication channel. Finally, in the third phase, the output text is constructed based on the keyword spotting and word confidence metrics. The confidence metrics indicate how much of the transcribed text is correctly converted from input speech based on the acoustic evidence [32,33].

#### 3.2. Translation of ISL Sentence from English Sentence

This section provides the details of the conversion process from English text to its corresponding ISL text. The words in the ISL sentence have been identified to generate corresponding sign movements. For the conversion from English to ISL, we use the Natural Language Toolkit (NLTK) [34]. The model of converting the ISL sentence from an English sentence is plotted in Figure 3. A detailed discussion of the translation process is presented below.



**Figure 3.** Model of ISL sentence generation from English sentence. PoS: Part-of-Speech; CFG: Context-Free Grammar.

### 3.2.1. Preprocessing of Input Text Using Regular Expression

If a user mistakenly enters an invalid/mispelled word, the “edit distance” function is used to obtain an equivalent valid word. A few examples of the misspelled words, along with the corresponding valid words, are presented in Table 2.

**Table 2.** Mapping of misspelled/invalid word into equivalent valid word.

Misspelled/Invalid Word	Equivalent Valid Word
“Hellllo”	“Hello”
“Halo”	“Hello”
“Hapyyyy”	“Happy”
“Happpppyyy”	“Happy”
“Noooooo”	“No”

The edit distance function takes two strings (source and target) and modifies the source string such that both source and target strings become equal. NLTK divides the English sentence into separate word–PoS pairs using the text tokenizer. The regular expression identifies the meaningful English sentence using the lexicon rule. During the preprocessing of input text, we define the regular expression (1) using the PoS tokens of the NLTK module. The regular expression starts with at least one verb phrase (VP) and is terminated with one noun phrase (NP). In the middle part, the regular expression can take zero or more number of any words that match PoS tokens (preposition (PP) or a pronoun (PRP) or adjective (JJ)). In a regular expression, + refers to one or more symbols, whereas \* refers to zero or more symbols. Therefore (VP)<sup>+</sup> represents one or more verb phrases. For example, our first sentence, “Come to my home”, starts with the verb phrase (“come”), followed by a preposition (“to”), pronoun (“my”), and ends with a noun phrase (“home”).

$$(VP)^+(PP|PRP|JJ)^*(NP) \tag{1}$$

where VP ∈ (VB,VBN), NP ∈ (NN), VB ∈ (“hello”, “Thank”, “Please”), VBN ∈ (“come”), PP ∈ (“to”, “with”), PRP ∈ (“my”, “you”, “me”), JJ ∈ (“Good”), NN ∈ (“home”, “morning”).

### 3.2.2. Syntactic Parsing and Logical Structure

After the preprocessing step, the NLTK module returns the parse tree based on the grammatical tokens (VP, PP, NP, etc.). Then, we construct the derivation tree of the Context-Free Grammar (CFG), which is similar to the parse tree of the NLTK module. CFG consists of variable/nonterminal symbols, terminal symbols, and a set of production rules. The nonterminal symbols generally appear on the left-hand side of the production rules, though they can also be introduced on the right-hand side. The terminal symbols appear on the right-hand side of the rules. The production rule generates the terminal string from the nonterminal symbol. The derivation tree can represent the derivation of the terminal string from the nonterminal symbol. In the derivation tree, terminal and non-terminal symbols refer to the leaves and intermediate nodes of the tree. Each meaningful ISL sentence has its own derivation tree. After the creation of the derivation tree, the leaves of the tree are combined to make a logical structure for the sign language. We have plotted the different derivation trees for a few sentences in Figure 4A–D.

#### Context-free grammar

$S \rightarrow VP PP NP | VP NP | VP VP PP NP$

$VP \rightarrow VB | VBN$

$PP \rightarrow \text{"to"} | \text{"with"}$

$NP \rightarrow PRP NN | JJ NN | PRP$

$VB \rightarrow \text{"hello"} | \text{"Thank"} | \text{"please"}$

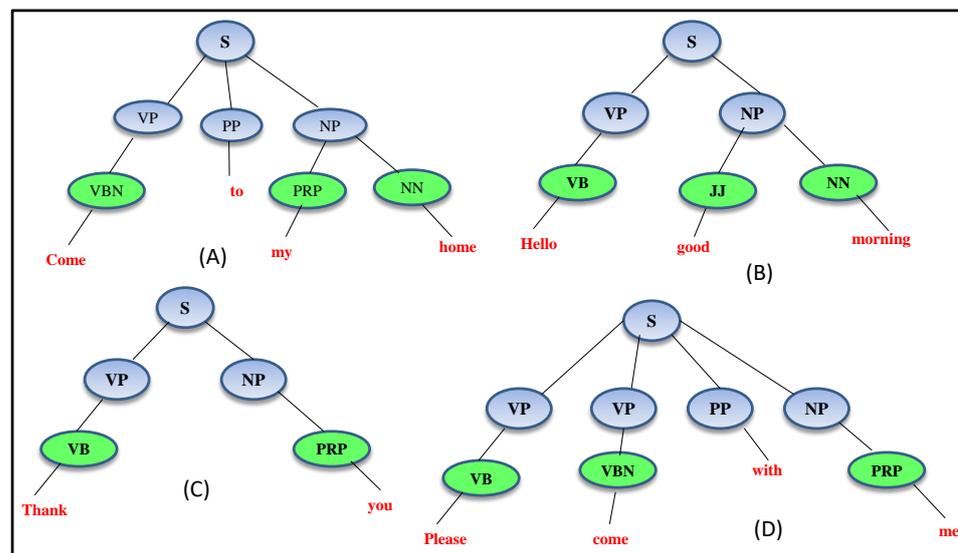
$VBN \rightarrow \text{"Come"}$

$PRP \rightarrow \text{"my"} | \text{"you"} | \text{"me"}$

$JJ \rightarrow \text{"Good"}$

$NN \rightarrow \text{"home"} | \text{"morning"}$

where  $\{\text{"hello"}, \text{"Thank"}, \text{"please"}, \text{"Come"}, \text{"my"}, \text{"you"}, \text{"me"}, \text{"Good"}, \text{"morning"}, \text{"to"}, \text{"with"}, \text{"home"}\} \in \text{terminals}$  and  $\{VP, PP, NP, VB, VBN, PRP, NN, JJ\} \in \text{nonterminals}$  of the context-free grammar.



**Figure 4.** Derivation tree for the sentences: (A) Come to my home, (B) Hello good morning, (C) Thank you, (D) Please come with me. Note: terminal symbols are represented in red, whereas green and blue refer to the nonterminal symbols of the derivation tree (generated from the above CFG). The start symbol (often represented as S) is a special nonterminal symbol of the grammar.

### 3.2.3. Script Generator and ISL Sentence

The script generator creates a script for generating an ISL sentence from the English sentence. The script takes a valid English sentence (after semantic parsing) as input and

generates the sequence tree, where each node of the tree is related to different gestures that are associated with the avatar movement. The sequence tree maintains the order of the motion performed by the avatar model.

The structures of the English and ISL sentences are quite different. The representation of ISL from the English sentences is done using Lexical Functional Grammar (LFG). The f-structure of LFG encodes the grammatical relation, like a subject, object, and tense of an input sentence. ISL follows the word order “Subject–Object–Verb”, whereas the English language follows the word order “Subject–Verb–Object” [35]. Moreover, the ISL sentence does not consider any conjunction and preposition in the sentence [36]. Some examples of mapping from English to ISL sentences are represented in Table 3.

**Table 3.** English sentence–ISL sentence mapping.

English Sentence	ISL Sentence
I have a pen.	I pen have.
The child is playing.	Child playing.
The woman is blind.	Woman blind.
It is cloudy outside.	Outside cloudy.
I see a dog.	I dog see.

### 3.3. Generation of Sign Movement

The generation of sign movements based on the input text is accomplished with the help of an animation tool called Blender [37]. The tool is popularly used for designing games, 3D animation, etc. The game logic and game object are the key components of the Blender game engine. We developed the 3D avatar by determining its geometric shape. The whole process for creating the avatar is divided into three steps. First, the skeleton and face of the avatar are created. In the second step, we define the viewpoint or orientation of the model. In the third step, we define the movement joints and facial expressions of the avatar. Next, we provide the sequence of frames that determine the movement of the avatar over the given sequence of words over time. Finally, motion (movement like walking, showing figures, moving hands, etc.) is defined by giving solid animation. The game engine was written from scratch in C++ as a mostly independent part and includes support for features such as Python scripting and OpenAI 3D sound. In this third module, we generate sign movements for the ISL sentence (generated in the second module). The entire framework of the movement generation of the avatar from the ISL sentence is described in Figure 5. For the generation of sign movement from ISL, initially, the animation parameters are extracted from the ISL sentence. Once the animation parameters are identified, the motion-list of each sign is performed using a 3D avatar model. In the proposed 3D avatar model, each movement is associated with several motions, and all such motions are listed in the “motionlist” file. A counter variable is initialized for tracking current motion. Each motion has its timestamps mentioning how long the action of different gestures will be performed. Each motion is generated by a specific motion actuator when some sensor event has occurred. The controller acts as a linker between the sensor and actuator. The conditional loop checks the maximum bounds of the number of motions, and it performs the motion one by one using a specific actuator (e.g., the actuator[counter]). The next valid motion is performed by incrementing the counter variable. If the value of the counter variable exceeds the number of motions in the “motionlist” file, then the counter variable along with the “motionlist” is reset to the default value, and the next movement will be performed.

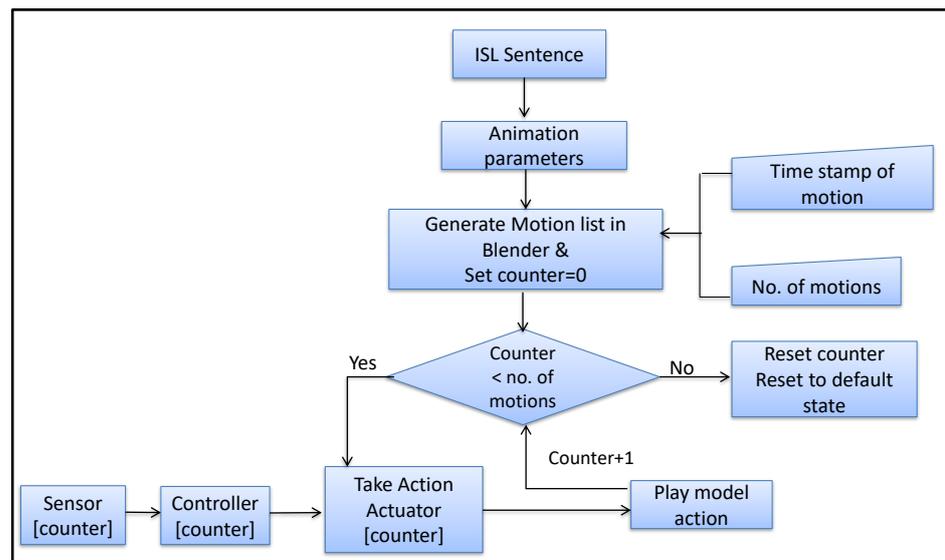


Figure 5. Movement generation of avatar from ISL sentence.

#### 4. Results

Here, we present a detailed analysis of the proposed sign language learning system using the 3D Avatar model. This section consists of four sections, namely: sign database (Section 4.1), speech recognition results (Section 4.2), results of translation process (Section 4.3), and generation of sign movement from ISL sentence (Section 4.4).

##### 4.1. Sign Database

We created a sign language database that contains sentences based on 50 daily used ISL words (e.g., I, my, come, home, welcome, sorry, rain, you, baby, wind, man, woman, etc.) and other dialogues between different users. We create 150 sentences that contain 763 different words, including the most used words in ISL. For each word, the sign movements were defined in the blender toolkit. The description of the sign database is depicted in Table 4. The vocabulary items were created based on the unique words in ISL. For better understanding, we represented four animation sequences of each word. For the sake of simplicity, we present some example sequences of sign movement (Table 5) using two animated series. The table shows the sign movements along with their actual words in English. All such sign movements were defined with the help of a sign language expert from a hearing-impaired school ('Anushruti') in the Indian Institute of Technology Roorkee.

Table 4. Dataset description.

Total Sentences	Total Words	Vocabulary	Running Words
150	763	365	50

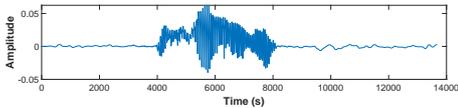
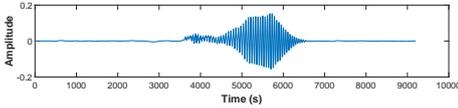
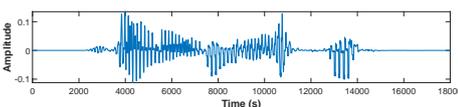
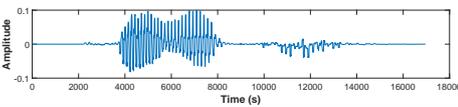
**Table 5.** English word to sign movements (each sign movement consists of 2 sequential motions).

English Word	Sign Movement	English Word	Sign Movement
Home		Night	
Morning		Work	
Welcome		Bye	
Rain		Baby	
Please		Sorry	

#### 4.2. Speech Recognition Results

The speech recognition is performed using the “IBM-Watson speech to text” service that converts English audio recordings or audio files into the respective text. The service takes a speech or audio file (.wav, .flac, or .mp3 format) with a different sampling frequency and converts the resulting text as output. The sampling frequency of our audio files is 16 KHz. The results of the speech recognition module for both isolated words (discrete speech) and complete sentences (continuous speech) are presented in Table 6. The X-axis denotes the time in seconds, and the Y-axis represents the amplitude of the signal. For the sake of simplicity, we take two discrete and two continuous speech signals for conversion.

**Table 6.** Mapping of the speech signal to text for different types of speech.

Speech Type	Speech Signal	Output Text (Using IBM-Watson Service)
Discrete		Hello
Discrete		Come
Continuous		Do you like it?
Continuous		Thank you

### 4.3. Results of Translation Process

This section illustrates the translation process of the proposed model. The translation process includes English to ISL sentence conversion and ISL sentence-to-sign representation. The evaluation of the proposed system was performed by dividing the generated sentences into a 80:20 ratio between the training and testing sets, respectively, and the Word Error Rate (WER) of the input word was recorded. The result of the text processing system is presented in Table 7, where the WER metric is derived from Levenshtein distance (edit distance function). Here we compare the word from the reference sentence and the output sentence. The distance calculates the number of edits/changes (insertion/deletion and modifications) required to convert the input text to the correct reference text. In Table 7, Ins, Del, and Sub refer to the number of insertion, deletion, and modification/substitution operations for converting source text to the proper target text, respectively.

**Table 7.** Text processing results based on Word Error Rate (WER).

WER (%)	Ins (%)	Del (%)	Sub (%)
25.2	3.3	7.1	14.8

For evaluating the performance of the translation system, some metrics have been considered: SER, BLEU, and NIST. SER computes the sign error rate during the generation of each sign from the ISL sentence. In this work, we recorded SERs of 10.50 on the test data. This error occurred due to WER happening during text entry input, which resulted into wrong sign generation by the avatar. BLEU and NIST are used for evaluating the quality of text during the translation from English (source language) to ISL (target language). The translation is done based on the multiple reference text (used from the vocabulary), and it calculates the precision score based on the unigram, bigram, . . . , *n*-gram model where *n* is the number of words in the reference text.

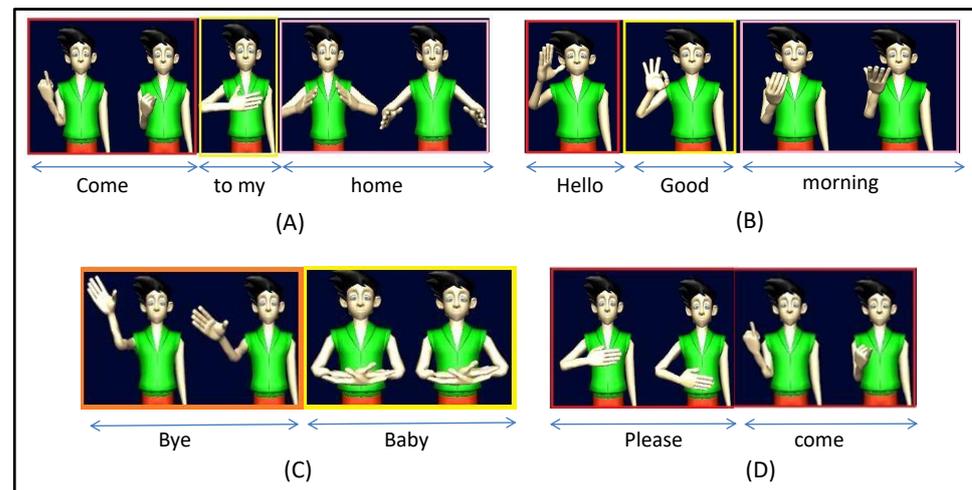
BLEU assigns equal weights to all *n*-grams, whereas NIST gives more importance to the rare words and small weights to the frequently used words in the sentence, so the overall score of NIST is better than BLEU. The result of the rule-based translation process is presented in Table 8. From Table 8, it can be concluded that the NIST score outperforms the BLEU score.

**Table 8.** Performance analysis of proposed translation system. SER: Sign Error Rate; BLEU: Bilingual Evaluation Understudy; NIST: National Institute of Standards and Technology.

SER	BLEU	NIST
10.50	82.30	86.80

#### 4.4. Generation of Sign Movement from ISL Sentence

After converting the ISL sentence from the English sentence in module 2 (Section 3.2), we proceeded to generate the sign movement for the ISL sentence. The avatar generates animated sign movements for each meaningful word. Here, we have used the avatar using blender software. We have plotted all the movements of ISL corresponding to English sentences. Figure 6A describes the actions (action 1, action 2, and action 3) of sign language representation for the English sentence “Come to my home”. Figure 6B depicts the sign language representation of the English sentence “Hello, Good morning” (action 1, action 2, and action 3), and Figure 6C describes the sign language representation of the English sentence “Bye baby” (action 1, action 2). Figure 6D represents the sign language representation (action 1, action 2) of the English sentence “Please come”.



**Figure 6.** Sign language representations of English sentences: (A) Come to my home, (B) Hello, good morning, (C) Bye baby, (D) Please come.

The quality of the proposed system was evaluated by adopting the Absolute Category Rating (ACR) [38] scheme. The ratings presented to the users are sorted by quality in decreasing order: Excellent, Good, Fair, Poor, and Bad. The performance is measured based on the output sign movements produced using various input speech or text. A majority rating of “Good” was recorded among the rater population of 25. A prototype video representation of the system was also made (available online: [https://www.youtube.com/watch?v=jTtRi8PG0cs&ab\\_channel=PradeepKumar](https://www.youtube.com/watch?v=jTtRi8PG0cs&ab_channel=PradeepKumar) (accessed on 22 December 2020)) on Youtube.

## 5. Conclusions

In this work, we developed a 3D avatar-based sign language learning system that converts the English speech or text into corresponding ISL movements. Initially, the input speech is converted into an equivalent English sentence using the IBM-Watson service. The converted English text is further translated into the corresponding ISL sentence using regular expression and script generator. Finally, each word of the ISL sentence is transformed into an equivalent sign movement, represented by a 3D avatar. The translation method (English–ISL and ISL–sign movement) has been evaluated by the SER, BLEU, and NIST metrics. The proposed model achieved an 82.30% BLEU score that represents

the translation accuracy from English to ISL sentences. The text-to-sign translation model (from ISL sentence to sign movement) achieved a 10.50 SER score, which signifies that 89.50% of signs were correctly generated by the 3D avatar model for the respective ISL sentence. It may be noted that the proposed system has been developed for a limited corpus, and no facial expressions were included, which can be considered as an important part of any sign language system. The transition between the signs while performing a sign sentence can be improved further by learning specific transitions based on hand positions while signing. Moreover, the sign language recognition system that converts a sign to text/speech is significantly more difficult to develop. Such a system can be added within the proposed framework to build a complete sign language interpretation system.

**Author Contributions:** Conceptualization, D.D.C. and P.K.; methodology, D.D.C., P.K. and S.M.; software, S.M.; validation, D.D.C. and P.K.; formal analysis, D.D.C. and P.K.; investigation, P.P.R. and M.I.; data curation, D.D.C. and P.K.; writing—D.D.C., P.K.; writing—review and editing, P.P.R. and M.I.; visualization, M.I.; supervision, P.P.R. and B.-G.K.; project administration, B.-G.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** No research funding has been received for this work.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kumar, P.; Roy, P.P.; Dogra, D.P. Independent bayesian classifier combination based sign language recognition using facial expression. *Inf. Sci.* **2018**, *428*, 30–48. [\[CrossRef\]](#)
2. Mittal, A.; Kumar, P.; Roy, P.P.; Balasubramanian, R.; Chaudhuri, B.B. A modified LSTM model for continuous sign language recognition using leap motion. *IEEE Sens. J.* **2019**, *19*, 7056–7063. [\[CrossRef\]](#)
3. Kumar, P.; Gauba, H.; Roy, P.P.; Dogra, D.P. A multimodal framework for sensor based sign language recognition. *Neurocomputing* **2017**, *259*, 21–38. [\[CrossRef\]](#)
4. Kumar, P. Sign Language Recognition Using Depth Sensors. Ph.D. Thesis, Indian Institute of Technology Roorkee, Roorkee, India, 2018.
5. Kumar, P.; Saini, R.; Behera, S.K.; Dogra, D.P.; Roy, P.P. Real-time recognition of sign language gestures and air-writing using leap motion. In Proceedings of the IEEE 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), Nagoya, Japan, 8–12 May 2017; pp. 157–160.
6. Krishnaraj, N.; Kavitha, M.; Jayasankar, T.; Kumar, K.V. A Glove based approach to recognize Indian Sign Languages. *Int. J. Recent Technol. Eng. IJRTE* **2019**, *7*, 1419–1425.
7. Zeshan, U.; Vasishta, M.N.; Sethna, M. Implementation of Indian Sign Language in educational settings. *Asia Pac. Disab. Rehabil. J.* **2005**, *16*, 16–40.
8. Dasgupta, T.; Basu, A. Prototype machine translation system from text-to-Indian sign language. In ACM Proceedings of the 13th International Conference on Intelligent User Interfaces, Gran Canaria, Spain, 13–16 January 2008; pp. 313–316.
9. Kishore, P.; Kumar, P.R. A video based Indian sign language recognition system (INSLR) using wavelet transform and fuzzy logic. *Int. J. Eng. Technol.* **2012**, *4*, 537. [\[CrossRef\]](#)
10. Goyal, L.; Goyal, V. Automatic translation of English text to Indian sign language synthetic animations. In Proceedings of the 13th International Conference on Natural Language Processing, Varanasi, India, 17–20 December 2016; pp. 144–153.
11. Al-Barahmtoshy, O.H.; Al-Barhamtoshy, H.M. Arabic Text-to-Sign Model from Automatic SR System. *Proc. Comput. Sci.* **2017**, *117*, 304–311. [\[CrossRef\]](#)
12. San-Segundo, R.; Barra, R.; Córdoba, R.; D'Haro, L.; Fernández, F.; Ferreiros, J.; Lucas, J.M.; Macías-Guarasa, J.; Montero, J.M.; Pardo, J.M. Speech to sign language translation system for Spanish. *Speech Commun.* **2008**, *50*, 1009–1020. [\[CrossRef\]](#)
13. López-Ludeña, V.; San-Segundo, R.; Morcillo, C.G.; López, J.C.; Muñoz, J.M.P. Increasing adaptability of a speech into sign language translation system. *Exp. Syst. Appl.* **2013**, *40*, 1312–1322. [\[CrossRef\]](#)
14. Halawani, S.M.; Zaitun, A. An avatar based translation system from arabic speech to arabic sign language for deaf people. *Int. J. Inf. Sci. Educ.* **2012**, *2*, 13–20.
15. Papadogiorgaki, M.; Grammalidis, N.; Tzovaras, D.; Strintzis, M.G. Text-to-sign language synthesis tool. In Proceedings of the IEEE 13th European Signal Processing Conference, Antalya, Turkey, 4–8 September 2005; pp. 1–4.
16. Elliott, R.; Glauert, J.R.; Kennaway, J.; Marshall, I.; Safar, E. Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Univers. Access Inf. Soc.* **2008**, *6*, 375–391. [\[CrossRef\]](#)

17. ELGHOUL, M.J.O. An avatar based approach for automatic interpretation of text to Sign language. *Chall. Assist. Technol. AAATE* **2007**, *20*, 266.
18. Loke, P.; Paranjpe, J.; Bhabal, S.; Kanere, K. Indian sign language converter system using an android app. In Proceedings of the IEEE 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 20–22 April 2017; Volume 2, pp. 436–439.
19. Nair, M.S.; Nimitha, A.; Idicula, S.M. Conversion of Malayalam text to Indian sign language using synthetic animation. In Proceedings of the IEEE 2016 International Conference on Next Generation Intelligent Systems (ICNGIS), Kottayam, India, 1–3 September 2016; pp. 1–4.
20. Vij, P.; Kumar, P. Mapping Hindi Text To Indian sign language with Extension Using Wordnet. In Proceedings of the International Conference on Advances in Information Communication Technology & Computing, Bikaner, India, 12–13 August 2016; p. 38.
21. Adamo-Villani, N.; Doublestein, J.; Martin, Z. The MathSigner: An interactive learning tool for American sign language. In Proceedings of the IEEE Eighth International Conference on Information Visualisation, London, UK, 16 July 2004; pp. 713–716.
22. Li, K.; Zhou, Z.; Lee, C.H. Sign transition modeling and a scalable solution to continuous sign language recognition for real-world applications. *ACM Trans. Access. Comput.* **2016**, *8*, 7. [[CrossRef](#)]
23. López-Ludeña, V.; González-Morcillo, C.; López, J.C.; Barra-Chicote, R.; Córdoba, R.; San-Segundo, R. Translating bus information into sign language for deaf people. *Eng. Appl. Artif. Intell.* **2014**, *32*, 258–269. [[CrossRef](#)]
24. Bouzid, Y.; Jemni, M. An Avatar based approach for automatically interpreting a sign language notation. In Proceedings of the IEEE 13th International Conference on Advanced Learning Technologies, Beijing, China, 15–18 July 2013; pp. 92–94.
25. Duarte, A.C. Cross-modal neural sign language translation. In Proceedings of the 27th ACM International Conference on Multimedia, Nice, France, 21–25 October 2019; pp. 1650–1654.
26. Patel, B.D.; Patel, H.B.; Khanvilkar, M.A.; Patel, N.R.; Akilan, T. ES2ISL: An advancement in speech to sign language translation using 3D avatar animator. In Proceedings of the 2020 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), London, ON, Canada, 30 August–2 September 2020; pp. 1–5.
27. Stoll, S.; Camgöz, N.C.; Hadfield, S.; Bowden, R. Text2Sign: Towards Sign Language Production Using Neural Machine Translation and Generative Adversarial Networks. *Int. J. Comput. Vis.* **2020**, *128*, 891–908. [[CrossRef](#)]
28. Kaur, K.; Kumar, P. HamNoSys to SiGML conversion system for sign language automation. *Proc. Comput. Sci.* **2016**, *89*, 794–803. [[CrossRef](#)]
29. Hanke, T. HamNoSys-representing sign language data in language resources and language processing contexts. *LREC* **2004**, *4*, 1–6.
30. Kennaway, R. Avatar-independent scripting for real-time gesture animation. *arXiv* **2015**, arXiv:1502.02961.
31. Zhang, Y.; Vogel, S.; Waibel, A. Interpreting bleu/nist scores: How much improvement do we need to have a better system? In Proceedings of the Fourth International Conference on Language Resources and Evaluation, Lisbon, Portugal, 26–28 May 2019; pp. 1650–1654.
32. Larson, M.; Jones, G.J. Spoken content retrieval: A survey of techniques and technologies. *Found. Trends Inf. Retrieval.* **2012**, *5*, 235–422. [[CrossRef](#)]
33. Lee, L.S.; Glass, J.; Lee, H.Y.; Chan, C.A. Spoken content retrieval—Beyond cascading speech recognition with text retrieval. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2015**, *23*, 1389–1420. [[CrossRef](#)]
34. Bird, S.; Klein, E.; Loper, E. *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*; O'Reilly Media Inc.: Newton, MA, USA, 2009.
35. Mehta, N.; Pai, S.; Singh, S. Automated 3D sign language caption generation for video. *Univers. Access Inf. Soc.* **2020**, *19*, 725–738. [[CrossRef](#)]
36. Zeshan, U. Indo-Pakistani Sign Language grammar: A typological outline. *Sign Lang. Stud.* **2003**, *3*, 157–212. [[CrossRef](#)]
37. Roosendaal, T.; Wartmann, C. *The Official Blender Game Kit: Interactive 3d for Artists*; No Starch Press: San Francisco, CA, USA, 2003.
38. Tominaga, T.; Hayashi, T.; Okamoto, J.; Takahashi, A. Performance comparisons of subjective quality assessment methods for mobile video. In Proceedings of the IEEE 2010 Second international workshop on quality of multimedia experience (QoMEX), Trondheim, Norway, 21–23 June 2010; pp. 82–87.